Evolving Communication through the Inference of Meaning

Andrew D. M. Smith
B.A. (Hons.), M.Sc.

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

to

Theoretical and Applied Linguistics,
School of Philosophy, Psychology and Language Sciences,
University of Edinburgh

September 2003

Abstract

In this thesis, I address the problem of how successful communication systems can emerge between agents who do not have innate or explicitly transferable meanings, cannot read the minds of their interlocutors, and are not provided with any feedback about the communication process. I develop a solution by focusing on the role of meanings within the framework of language evolution, and on communication through the repeated inference of meaning.

Much recent work on the evolution of language has concentrated on the emergence of compositional syntax as the crucial event which marked the genesis of language; all the experimental models which purport to demonstrate the emergence of syntax, however, rely on models of communication in which the signals are redundant and which contain pre-defined, structured meaning systems which provide an explicit blueprint against which the syntactic structure is built. Moreover, the vast majority of such meaning systems are truly semantic in name only, lacking even the basic semantic characteristics of sense and reference, and the agents must rely on mind-reading or feedback (or both) in order to learn how to communicate.

By contrast, at the heart of this thesis is a solution to the signal redundancy paradox based on the inference of meaning and the disambiguation of potential referents through exposure in multiple contexts. I describe computational models of meaning creation in which agents independently develop individual conceptual structures based on their own experiences of the environment, and show through experimental simulations that the agents can use their own individual meanings to communicate with each other about items in their environment. I demonstrate that the development of successful communication depends to a large extent on the synchronisation of the agents' conceptual structures, and that such synchronisation is significantly more likely to occur when the agents use an intelligent meaning creation strategy which can exploit the structure in the information in the environment.

Motivated by research into the acquisition of language by children, I go on to explore how the introduction of specific cognitive and lexical biases affects the level of communicative success. I show that if the agents are guided by an assumption of mutual exclusivity in word meanings, they do not need to have such high levels of meaning similarity, and can instead communicate successfully despite having very divergent conceptual structures.

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Andrew D. M. Smith

Acknowledgements

The acknowledgements section: what does it provide, and to whom?

- To the author,

  1. A rare and invaluable opportunity to document, (relatively) publicly and in perpetuity, with the freedom both to be outrageously ostentatious and cringingly obsequious, and to indulge in fantastically obscure, pretentious and grandiloquent linguistic constructions, such as the pointless yet enchanting preposition of postpositional phrases, his multiform and heartfelt gratitude to those inspirational collaborators, supportive friends and casual acquaintances who have in some way aided and abetted the germination, gestation and final efflorescence of his *magnum opus*;
  2. Furthermore, the space to fulfil the tacitly understood and solemn duty of recording in reciprocation the appellations of all those in whose own theses' acknowledgements sections the author's name has already previously been logged, and, in anticipatory reciprocation of similar inscriptions, the appellations of all those in whose theses' acknowledgements sections the author might wish his name to be documented subsequently.

- To the reader, an easy and painless way of

  1. checking whether their name has been included;
  2. appearing, to the casual observer, that they are engrossed by the subject matter of the thesis.

The aforementioned obfuscatory circumlocutions notwithstanding, it behoves me to record my sincere thanks to those people and organisations without whom this thesis would never have been completed.

Firstly, to my supervisors Jim Hurford and Simon Kirby, especially to Jim's indefatigable enthusiasm, helpful and insightful criticism, and vital encouragement during the last four years.

To my office mates Kenny Smith and Henry Brighton, for the benefit of their wisdom, for their constant cheerfulness, and of course for their unsurpassed expertise in predicting football results.

To all the other members of the Language Evolution and Computational Research Unit over the time I've been here, namely: Hajime Yamauchi, Joseph Poulshock, Paul Vogt, Jean-Louis Dessalles, Takashi Hashimoto, Jelle Zuidema, Anna Parker, Viktor Trón and Tim O'Donnell, for the welcoming and enjoyable environment which they have helped provide.

Financially, to the Economic and Social Research Council, who initially took a chance on an Information Manager working in cancer epidemiology to award me a PhD studentship four years ago, and to Bob Ladd and the department of Theoretical and Applied Linguistics, who also took a chance by giving me a temporary lectureship in linguistics which enabled me to finish.

In computing terms, to Cedric MacMartin and the rest of the linguistics computing support team, who provided the ideal facilities to undertake the simulations.

Most importantly, to my wife Caroline, for initially encouraging me to pursue the idea of a PhD, for her crucial support throughout, and for her unwavering and enduring love. For everything, really.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

> In the study of language, we embrace the very definition of what it means to be human ... (Locke, 1993, p. 4)

1.1 Universality and Diversity

The capacity for using language is not only unique to humans, but is also fundamental to our understanding of what it means to be human. But what is it that speakers of a language actually know, and how is this knowledge represented mentally? An important way of studying such issues is through examining the development of children, exploring how they develop an understanding of the world around them and acquire the language of the community in which they are raised.

Children, virtually without exception, acquire language at a very early age; the system of language they acquire is extremely complicated, and yet they acquire it rapidly, with few (and relatively predictable) errors, without being taught and with only limited experience of it. Clearly, to some extent, we are all genetically programmed for language. On the other hand, the *specific* languages the child acquires are those languages which they hear spoken by the people they interact with; equally clearly, there are enormous differences between the languages of the world. Indeed, there are generally reckoned to be between 6,000 and 7,000 different languages in the world today (Nettle, 1999; Song, 2001), although even this could be an underestimate, depending on where the line between a language and a dialect is drawn. How can we reconcile the universality of language as a general, distinctively human phenomenon with the diversity of languages seen around the world?

The *nativist* theory of language, initially developed by Chomsky (1965), is concerned with describing a person's *internal* knowledge of their language (I-language), or the mental instantiation of language; this is distinguished from a person's actual *external* use of the language (E-language), or the behaviour produced by the user in response to a particular set of circumstances. In order to account both for the universality and relative ease with which language is acquired, in notable contrast to the non-universality and relative difficulty in the mastering of other cognitive tasks such as playing music, nativists suggest that all children are innately specified with a domain-specific[1] capacity to acquire grammar, called Universal Grammar (UG). There are a number of powerful arguments in favour of a universal, innate blueprint for grammar acquisition put forward by nativists like Chomsky (1965), Wexler (1991), Pinker (1994), Lightfoot (1999), the most frequently promoted of which is known as the *poverty of the stimulus*; this runs, essentially, as follows:

1. A child is exposed to a set of primary linguistic data (PLD) when it is acquiring language; this data
   - is quantitatively finite;
   - is qualitatively relatively inaccurate, containing numerous errors (like slips of the tongue) in comparison with the I-language which generated it;
   - contains only positive examples, so the child receives no evidence of sentences which are not part of the language.
2. All human languages are infinitely expressive, so the child must generalise from this finite set of data to an infinite set of sentences, and so learn to produce and understand sentences which it never hears.
3. There are an infinite number of possible languages logically consistent with the PLD; increasing the number of sentences does not reduce the set of possible languages (Gold, 1967).
4. Only negative evidence will allow a reduction in the set of possible languages, but negative evidence does not occur very often, if at all (Bowerman, 1988).
5. Despite this insufficient evidence, the child generalises to its mother tongue, (or more accurately to a very close approximation of its mother tongue which is comprehensible to other speakers).

[1]Domain-specificity, the idea that this capacity to acquire grammar is specific to language and cannot be used for any other task, is contrasted with domain-general cognitive processes, which can be used for different tasks across many domains.

6. Therefore, there must be some constraints on what the child can learn, some "innate ideas and principles of various kinds that determine the form of the acquired knowledge in what may be a rather restricted and highly organised way" (Chomsky, 1965, p. 48); the child's strategy for creating their internal grammar is known as the Language Acquisition Device (LAD).

Under the nativist paradigm, language acquisition can succeed despite the lack of sufficient input, because Universal Grammar constrains the hypothesis space from which the child chooses. The variation among actual existing languages can be redefined as making choices from within this hypothesis space of possible human languages, or, as it is often expressed, setting a finite, innate set of parameters through cultural interaction within their particular community as they acquire language. For instance, if children are exposed to a language like Swahili, in which objects generally follow their verbs and adjectives follows nouns, they automatically flick the Head-Ordering parameter to the head-first setting. On the other hand, if they hear the verbs following objects and nouns following adjectives, as in Japanese, they flick the parameter to the head-last setting (Pinker, 1994). Linguistic research for nativists can now be centred on discovering the number and structure of these innately-specified parameters, the default and possible values which they can take, and the triggers which enable them to be set during language acquisition.

There are, however, difficulties with this approach, which are not helped by the fact that there is no consensus whatsoever on how many switchable parameters exist, or even that the principles and parameter thesis is broadly correct; there are indeed numerous different competitor linguistic theories, including, in by no means an exhaustive list: Categorial Grammar (Steedman, 2000); Cognitive Grammar (Langacker, 1987); Head-Driven Phrase Structure Grammar (Sag & Wasow, 1999); Lexical-Functional Grammar (Bresnan, 2001); Radical Construction Grammar (Croft, 2001), Role and Reference Grammar (Van Valin & LaPolla, 1997); and Word Grammar (Hudson, 1984, 1995). Moreover, despite the attractiveness and simplicity of the triggering parameter-setting account of language acquisition, little attempt is made to address the very pertinent questions of how a child hearing an unfamiliar set of words knows which word is the object and which the verb, what a linguistic head is, or, on an even more basic level, what verbs and nouns are. In addition, the 'all-or-nothing' approach of changing parameters based on trigger sentences seems to be clearly at odds with empirical results in the field of language development, in which the expected discrete changes in behaviour when a parameter is switched are simply not seen (Tomasello, 2001a).

In contrast to the domain-specificity favoured by nativists, *empiricists* assert that language acquisition can be explained through *domain-general* processes. One of the most prominent and vociferous amongst those who reject the Chomskyan approach is Sampson (1997), who asserts that "biological constraints on language are limited to matters which are 'trivial' because they follow from properties of our speech and sense organs ... " (Sampson, 1997, p. 25). Sampson systematically attacks each of the nativist arguments, including the poverty of the stimulus; he finds that, contrary to assertions made by nativists, the PLD available to children:

- contains remarkably few disfluencies (in fact, almost none at all) (Newport, Gleitman, & Gleitman, 1977);
- does contain evidence which allows children to rule out certain hypotheses (Pullum & Scholz, 2002);

It is unarguable, however, that the input to the child is finite, and that from this it must generalise to an infinite set, but this too is not the insurmountable problem it might appear. Connectionist networks, for instance, have been shown to induce patterns from irregular input, and to generalise these patterns to novel information, as long as the network focuses on simple sentences first, which provide it with information about categories and agreement which it can use to learn the more complex sentences (Elman, 1993).

1.2 The Evolution of Language

The transition between using short, finite communication systems to the capacity for an infinitely expressive language is, according to Maynard Smith and Szathmáry (1995), the most recent major transition in the evolutionary history of life on earth, but what this transition involved is still an open question.

Language as an Organ

In their seminal article which re-ignited much of the recent burgeoning interest in language evolution, Pinker and Bloom (1990) argue persuasively that "a specialization for grammar evolved by a conventional neo-Darwinian process"(Pinker & Bloom, 1990, p.707), suggesting that humans have evolved an innate, genetically specified module in the brain, which specifies a formal coding of the principles of Universal Grammar. In this way, language is embodied like any other bodily organ, while still simultaneously being assumed to be somehow resident within the brain, with its precise position and constitution unresolved; more importantly, it is taken, in accordance with the nativist view outlined above, to be specifically tailored to the acquisition and maintenance of language.

But if we accept that language is expressed through the genes, how did this happen? Pinker and Bloom (1990) are firmly of the opinion that the selective advantage of the communicative function of language can explain the evolution of the language faculty itself:

> "Language shows signs of complex design for the communication of propositional structures, and the only explanation for the origin of organs with complex design is the process of natural selection." (Pinker & Bloom, 1990, p.726)

They argue that although no single mutation could have led to an entire universal grammar, a parent with a primitive grammar G could have given birth to a mutant offspring with a slightly more enhanced grammar G', and that such a process could have occurred in repeated increments until the Universal Grammar humans have today was reached. If we do accept the existence of a complex language organ, it does seem irresistible to agree that natural selection must have produced it, and "that the LAD evolved as an adaptation to acquisition should be our null hypothesis" (Kirby, 1999, p. 124), although we are still a long way from explaining the conditions which led to its appearance only in humans. Jackendoff (2002) has recently tried to flesh out Pinker and Bloom's position of incremental evolution in much more detail, by putting forward an ordered set of steps from primate conceptual structure to modern language through the use of symbols out of context, the availability of an unlimited vocabulary, combinations and concatenations of sounds and symbols, hierarchical phrase structure, abstract semantic relationships, the emergence of grammatical categories, and finally inflectional morphosyntax.

On the other hand, Chomsky (1988), perhaps somewhat surprisingly given his introduction of the very idea of Universal Grammar, argues, as does Lightfoot (1999), that the role of natural selection in language evolution is very limited, and that the parts of the brain necessary for language, despite their supposed linguistic domain-specificity, were reappropriated (or *exapted*) by language after having evolved for a separate, unspecified, cognitive purpose. Others have used similar arguments to argue against a language-specific learning device itself, arguing that particular physical and cognitive characteristics were selected for, and that the combination of these somehow kick-started language:

- Dunbar (1996) suggests that language evolved into a sort of verbal grooming, as a means of maintaining the same system of reciprocal altruism, as the human group size increased past the point at which all members could be physically groomed.

- Deacon (1997) argues convincingly that the construction of abstract semantic models, where symbolic representations are linked to other symbolic representations, made the cognitive breakthrough which allowed language to follow. Crucially, he also sees the development of language as instrumental in the continued shaping of the brain, as they evolved together, each re-inforcing the development of the other.

- Bickerton (1998) argues for a single mutation, one which created a connection in the brain between the social intelligence needed for a system of reciprocal altruism (such as that found in apes, and maintained by grooming), and a primitive protolanguage, and which "led directly to a cascade of consequences that would, in one rapid and continuous sequence, have transformed protolanguage into language substantially as we know it today" (Bickerton, 1998, p.353).

- Carstairs-McCarthy (1999) proposes that language is a "by-product of a change in the anatomy of the vocal tract", coupled with an "expectation that different vocalisations should mean different things" (Carstairs-McCarthy, 1999, p.226).

Language As An Organism

An alternative view to the nativist hypothesis focuses not on the biological manifestation of grammatical rules in a language organ, but instead on linguistic structures themselves adapting to fit the brain. This approach has been put forward using various appealing metaphors: *language as an organism* (Christiansen, 1994); *language as a virus*, with its users as hosts (Deacon, 1997); *utterances competing for selection* (Croft, 2000). According to all these accounts, it is useful to focus on the fact that languages are made up of utterances, and that these utterances are being repeatedly used by speakers and understood and recognised by hearers. The two distinct manifestations of language which we recognise from the Chomskyan account are reconfigured as distinct phases in the life cycle of the language: an internal grammar (I-language) forms the basis for the language's expression (E-language), which forms the basis for its subsequent recognition and reanalysis into another internal grammar (I-language); because of this continual cycle of expression and re-interpretation, languages can evolve culturally, as well as genetically, and, other things being equal (such as the relative health of the language speakers, the size of the population etc.), those languages which can be readily re-interpreted by their hosts are more likely to survive than those which cannot be.

Human languages, in this paradigm, have adapted to fit human cognitive structure, because utterances which are well matched to this structure have thriven, while those which are not suited have not persisted. An example of a part of an utterance which has not, in general, persisted, is the past participle 'thriven' used in the last sentence; this form has, through its infrequent use, been all but replaced by the analogically regular 'thrived'. Computational models which explain the maintenance of irregularities in terms of their increased frequency of use, while less-frequently used words must be systematised into a regular paradigm have been proposed by Hurford (2000), Kirby (2001) and others, while Kirby (1999) has also explored in detail how both parametric and hierarchical language universals[2] described by Greenberg (1966) can be explained elegantly by focusing on how processing complexity affects the transmission of language between speaker and hearer. In a similar fashion, Brighton (2002) shows how compositional syntax is likely to emerge under the specific circumstances of a complex meaning space structure and the poverty of the stimulus. Brighton suggests that the poverty of the stimulus, rather than implying the existence of the LAD, as in the Chomskyan position, is on the contrary actually a necessary *pre-condition* for the emergence of complex language, or, as Zuidema (2003) expresses it: "the poverty of the stimulus solves the poverty of the stimulus."

1.3 Computer Simulations of Language Evolution

Until recently, nearly all theories on the evolution of language have been based on the intuitions of their authors, and whilst they have been supported by evidence from a wide range of fields, they are, to a great extent, essentially unfalsifiable. An alternative strategy, and the one which I use in this thesis, is to build a model of a particular real-world phenomenon, which will allow the validation or contradiction of these theoretical predictions. Simulations can be run using the model, and if the behaviour exhibited by the simulation resembles, to a sufficient extent, the behaviour of the natural phenomenon, then the model on which the simulation is built can be advanced as a possible explanation for the phenomenon. Moreover, at a deeper level, the systems we are trying to model are dynamic and complex, with many interactions between variables at different levels; computational simulations explicitly allow experimenters to probe these different aspects of a complex system, to discover which factors in the model are the important explanatory factors which the phenomenon depends on, and which are unimportant for a particular issue. In particular, intuitive philosophical predictions turn out to be notoriously error-prone in this kind of theoretical thought experiment. In this vein, there has

[2]This technical distinction is made between *parametric* universals, in which a language having property x implies also having property y, and having property y implies having property x, and *hierarchical* universals, in which having property x implies having property y, *but not necessarily vice versa.*

recently been much effort expended to develop models of the evolution of aspects of human language. In particular, computational models have been developed which shed light on the evolution of an innate LAD (Yamauchi, 2001; Turkel, 2002), the emergence of phonological systems (de Boer, 2001, 2002), lexical and vocabulary development (Hurford, 1989; Vogt, 2000; Steels & Kaplan, 2002), conceptual development (Steels, 1996b; de Jong, 2000; Belpaeme, 2002), the emergence of syntax (Kirby, 2000, 2001; Briscoe, 2002), and, more specifically, compositionality (Batali, 2002; Brighton, 2002, 2003; Kirby, 2002; K. Smith, 2002a, 2003).

One of the major focuses of work in this field has been, as mentioned above, the evolution of syntactic structure, on the grounds that it is the crucial event which marks both the genesis of language and the defining criterion which separates it from animal communication systems. Kirby (2002), for example, demonstrates that syntax can arise from unstructured communication systems through the simple ability to create general rules based on coincidental correspondences between parts of utterances and parts of meanings; the general rules can generalise beyond their input, can generate more utterances than idiosyncratic rules, and so are replicated in greater numbers in future generations. A similar account is provided by Batali (2002), whose agents hypothesise mappings between strings and meanings on the basis of exemplars. Batali's agents are endowed with the ability to combine and modify phrases, rather than the ability to generalise rules, but again advantage is taken of coincidental correspondences between utterances and meanings to develop syntax. These accounts of how syntax emerged are given theoretic credence by Wray (1998, 2000), who argues that holistic expressions used in ritualised social situations would have been reanalysed as having compositional semantics, leading to syntax via a generalisation mechanism similar to that described by Kirby (2002). In these accounts, language can clearly be seen as a dynamic, self-organising system (Steels, 1996c), within defined parameters such as a tendency to generalise or to find structure in phrases.

Despite these exciting findings, however, there are some major problems with the assumptions behind simulations such as these, which the model I describe in this thesis seeks to overcome. Firstly, the 'emergent' syntax develops only because the utterances in the simulations are explicitly coupled with pre-existing, structured semantic representations. These semantic representations are already compositional and recursive, and the agents are endowed with a symbolic grammar, so in retrospect it is no great surprise that the syntax produced by the agents also turn out to be compositional and recursive, stored symbolically, and essentially a replica of the semantic representation, as Nehaniv (2000) and others have argued. Explanations of the origin of such meanings, and of how they become associated with the signals, are by contrast a major focus of this thesis.

Secondly, the meanings in such simulations are invariably part of the linguistic transfer between the two communicating agents; as well as hearing an utterance, the agent is given the meaning to which it corresponds, before it analyses the (coincidental) correspondences. This design feature of the simulations ignores one of the most baffling, and important features in language acquisition: meanings are clearly *not* explicitly transferred between speaker and hearer, and yet children *do* manage to derive the meanings; this paradox of signal redundancy, where the transfer of the meaning makes the use of the signal redundant, is explicitly avoided by my model.

In addition, attempts to develop learnt communication systems frequently involve some sort of reinforcement learning process (Steels, 1999; de Jong, 2000), which has the primary role in guiding the learning mechanism. Oliphant (1999) points out, however, that such error signals, which work well on an evolutionary timescale, are less useful over an individual's lifetime where failure to communicate might mean immediate death, and indeed even the very existence of reliable error signals is questioned by many authors on child language acquisition (Bloom, 2000). A further aim of this thesis is to explore the conditions under which communication can emerge without the need for error signals and feedback.

A major guiding principle behind this thesis is that semantic complexity is a pre-requisite for the emergence of syntax; indeed it has been hypothesised that the need to communicate semantically complex propositions has itself been the driving force behind the development of syntax (Schoenemann, 1999). I argue that the construction of meanings, and then learning which of these meanings are relevant, are fundamental parts of the language development process which cannot be overlooked or assumed in investigations into language evolution.

1.4 Outline of Thesis

In this thesis, therefore, I will present a solution to the problem of the development of successful communication systems which rely on neither innate nor explicitly transferable meanings, neither on the agents being able to read their interlocutors' minds, nor on them receiving feedback about the meaning creation and communication processes, by focusing on the role of meanings themselves within the framework of language evolution, and on communication through the inference of meaning.

Chapter 2: I start by exploring the philosophical nature of meaning, and in particular the difficulty of describing the meanings of words. Four important issues are dealt with in this chapter: the sense relationships between meanings; their mental representation; their referential grounding; and their acquisition. I explore a number of the general semantic relationships between meanings which I make use of in the model described later in the thesis, and investigate competing arguments concerning the mental representation of meanings as categories in the brain. Fundamentally, meanings must be grounded in reality by referring to objects and events in the environment, although this is seldom acknowledged in language evolution models; given this, I discuss different theoretical models by which grounded concepts can be acquired.

Chapter 3: I then confront the crucial problem of lexical acquisition, of how the meanings of words are learnt, starting with the philosophical paradox of the indeterminacy of meaning, then moving on to a detailed discussion of the various cognitive biases which have been proposed by developmental psychologists and linguists to get round this paradox while still assuming that meaning is inferred from somewhere; in particular, I discuss the anti-synonymy biases of the Principle of Contrast and the Mutual Exclusivity Assumption which are then explicitly encoded in the simulations described in chapter 9.

Chapter 4: The work described in chapters 2 and 3 is then brought together by investigating in detail the semantic nature of recent simulations of language evolution, both in terms of the representation and creation of meanings. I show that, although a semantic realm is essential to these experiments as a blueprint on which an emergent syntactic system can be parasitic, the meaning systems therein lack the most basic ingredients necessary and are actually semantic in name only. From the review of the meaning representation and creation processes, I justify the model which I will use in the experiments in later chapters.

Chapter 5: I describe my model of meaning creation, showing how agents can develop individual and divergent conceptual structures which are yet grounded in their experiences, and exploring the properties of the model in terms of conceptual development and the adequacy of the description of the world which it provides.

Chapter 6: The model is then extended to explore communication between agents under various conditions. Firstly, I discuss communication in the abstract, showing that the division between public and private knowledge is fundamental to a model which tries to avoid the signal redundancy paradox. I show the importance of lexical bidirectionality and of accommodating the hearer in building successful communication systems. The model is then brought together with a description of the lexicon at the heart of the strategy I adopt, which I call *introspective obverter*.

Chapter 7: The communication model is then tested thoroughly, and I show that agents' inference of meaning from context produces remarkably high levels of communicative success. I then explore the relationship between conceptual similarity and communicative success, and explain why randomly generated conceptual structures are so similar to each other in practice. Finally, the Gricean nature of the agents' communicative procedure is explored, explaining why the levels of communicative success in the simulations are regularly higher than the levels of meaning similarity.

Chapter 8: In this and the following chapter, I explore both the meaning creation and communicative procedures in much more detail, with comprehensive computational experiments to explore the impact of cognitive biases and the specific environmental pressures which allow the agents to ground their conceptual structures. I show that an intelligent meaning creation strategy can produce very successful communication, but its impact is primarily reserved for experiments when the agents live in a structured world, which the intelligent strategy can exploit to build relevant conceptual structures, which themselves lead to high rates of communicative success.

Chapter 9: Finally, I then link the lexical acquisition process back to the communicative biases discussed in chapter 3, implementing the *mutual exclusivity assumption* in the model, to explore its impact on both meaning similarity and communicative success. I show that the assumption of mutual exclusivity results in consistently high levels of communicative success, particularly in a structured world.

CHAPTER 2

Meanings

> "Man possesses the ability to construct languages capable of expressing every sense, without having any idea how each word has meaning or what its meaning is . . . " (Wittgenstein, 1921/2001, section 4.002)

2.1 Introduction

At first glance, Wittgenstein's assertion that humans use words to express meanings, but yet have no knowledge of what the meaning of a word is, seems strangely counter-intuitive. Without knowledge of what words mean, no manipulation of them appears possible, and it appears that successful communication would be impossible and trying to communicate would be pointless. Indeed, is it really communication at all, if the speaker of a sentence does not know the meaning of what he says? To take an analogy with another field of study, could we be said to be doing mathematics, if we had no knowledge of what the mathematical symbols we used stood for, and what they signified?

On the other hand, if we do try to define the sense of a word, the heart of its meaning, we quickly find ourselves in another kind of paradox, because we must rely on using other words to describe and explain the meanings of the original words which we are trying to define. We use word forms to communicate, to express meanings, because meanings appear to be, of themselves, inexpressible. Word forms can be uttered, but they themselves do not have any meaning of themselves. Meanings and word forms operate in different mediums of thought and expression, and the links between them, which underpin language itself, are not only arbitrary, as Saussure (1916) pointed out, but supported only by social convention and repeated use.

It is clear that the sense of a word can be defined by its relationship to the senses of other words, and that understanding these relationships holds the key to working out what the words themselves mean. Many such relationships have been identified by semanticists (Lyons, 1977; Cruse, 1986); in section 2.2, I will briefly discuss a few of the most important examples, which I will make use of in my model of unguided meaning creation which is described in detail in chapter 4. In section 2.3, I explore the problem of categorisation, and the heated debate around the nature of the mental representation of categories. In section 2.4, I look at grounding meanings, and at how innate animal communication systems can come to be shaped by natural selection, before moving on in section 2.5 to theories of human concept acquisition, both innatist and empiricist.

2.2 The Nature of Meaning

2.2.1 Semantic Relationships

Hyponymy

Hyponymy describes the relationship between one word and another word which has a more general meaning. For example, if x is a cat, then x is necessarily also a mammal, and so the word CAT is a hyponym of MAMMAL, and conversely, MAMMAL is a superordinate of CAT. Hyponymy has two interesting features which have implications for how meaning may be structured. Firstly, it is by its definition in terms of logical implication non-symmetrical, so that the converse of the logical statement above (if x is a mammal, then x is necessarily a cat) does *not* hold. Secondly, it is also transitive, so that if one word is the hyponym of a second, which is itself the hyponym of a third, then the first word is necessarily also a hyponym of the third word. We can see this by considering that if x is a tabby, then x is a cat. Taken together with our original statement, we know that if x is a tabby, then x is also a mammal.

Hyponymy, therefore, being both a non-symmetrical and transitive relationship between words, introduces the notion of hierarchy into our model of meaning, with general terms at the root of a tree which branches out into many more specific hyponyms, which themselves can have further hyponyms, as in figure 2.1. This notion of meaning being structured in a hierarchical fashion which is represented dendritically provides us with a relatively straightforward, yet powerful way to visualise a potentially infinite number of meanings from any one particular meaning, and which I shall use in chapter 5 to create

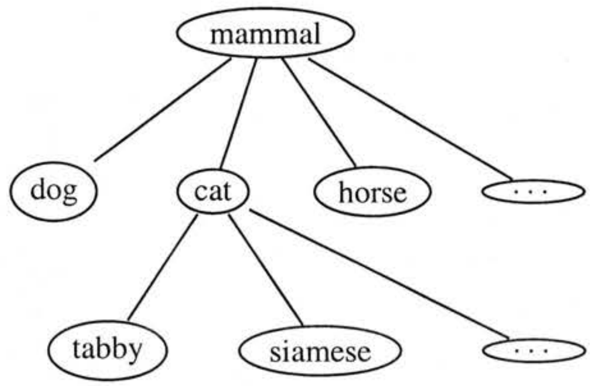


Figure 2.1: Part of a hierarchical model of meaning based on hyponymy.

a flexible system for dynamic meaning creation, in which I will make use of the notion of hyponymy in terms of *generalisation* (moving towards the root of the tree) and *specialisation* (moving towards the leaves of the tree).

Antonymy

Antonymy describes the relationship between a word and its opposite, and appears to be one of the most basic of semantic relationships. We may not be too surprised to note, however, that there are many types of opposition, with subtle differences in the way they behave. *Gradable* antonyms, such as WET/DRY, HOT/COLD are words which express meanings on some sort of relative scale, whereas *ungradable* binary antonyms, such as ALIVE/DEAD, MALE/FEMALE, express complementary propositions, which entail the negation of their opposite proposition. For example, a hot day in Scotland is likely to be considerably cooler in objective terms than a hot day in Tanzania, yet both are still relatively hotter than their respective cold days. On the other hand, a dead cat is just as dead in Europe as it is in Africa, and a cat being dead always implies that it is not alive. A word is the *converse* of another if they both refer to the same relationship between two entities, but the nature of the relationship is reversed, so if x is above y, then y is necessarily below x.

The important feature which unites all these different relationships under antonymy is that of dichotomy or binary opposition, of dividing the world up into two complementary meanings. For our purposes, it is not important whether these complementary meanings are absolute or relative, only that they divide things into two groups.

> "[B]inary opposition is one of the most important principles governing the structure of languages; and the most evident manifestation of this principle, as far as the vocabulary is concerned, is antonymy." (Lyons, 1977, p.271)

If we apply the notion of dichotomy to the hierarchical tree structure described above which can represent hyponymy, we can develop a tree structure in which every node has two co-hyponyms, each of which is the antonym of the other[1]. A binary branching structure is preferred because of the importance of binary opposition in the organisation of language, and because it is simpler both conceptually and computationally than a tree structure with unlimited branching, yet is still infinitely expressive and contains a straightforward representation of both hyponymy and antonymy. This characteristic makes it potentially suitable for an abstract, computational model of meaning, provided that we recognise that binary branching is not meant to be an exact model of how meaning is actually structured[2].

Synonymy

Synonymy is the simplest of relationships between senses, being the identity relationship, where the two senses are the same[3]. Despite the simplicity of the idea behind this relationship, and despite the fact that speakers can readily produce examples of synonymy in their language, even the existence of natural synonymy is controversial.

The problem is in the identity relation itself, and how rigidly we want to hold to its definition. If we look for hard synonyms, or words which have the same sense as each other and are interchangeable in all contexts, then it is extremely difficult to find suitable candidates. For instance, although in English both 'freedom' (of Germanic origin) and 'liberty' (of Romance origin) appear to have the same sense, and it is just an accident of history that both are extant in the modern language, they are not truly interchangeable. One man's terrorist could not usually be described as anyone's liberty fighter. Of course, 'freedom' and 'liberty' are indeed soft synonyms which have the same sense in the vast majority of cases, but they are not always interchangeable. Clark (1987), indeed, argues persuasively that there are no synonyms at all in natural languages, and that every

[1]In reality, we cannot say that all co-hyponyms are antonyms of each other, as we can readily produce trees such as DISABLED, with the hyponyms DEAF, BLIND, etc. which are not mutually exclusive.

[2]Although the tree in figure 2.1 can be converted into a binary branching tree, it can only be done so by creating many strange categories such as UNDOG, UNCAT. This, however, does not affect the expressivity of the binary tree, only the plausibility of some of the categories contained within it.

[3]Logically, we can consider it a special case of hyponymy, where each term is a hyponym of the other, and so *both* logical statements "if x is a y, then x is necessarily a z" and "if x is a z, then x is necessarily a y" are true.

pair of potential or apparent synonyms reflects a contrast in terms of dialect, register or connotation.

Returning to the hierarchical tree structure which frugally represents hyponymy and antonymy, it is straightforward to define synonymy tightly using this system: two words are synonyms if they correspond to the same point on a particular hierarchical meaning structure.

2.3 Categorisation

The properties which characterise objects are the foundations on which categorisation is built. In the real world, there are many different suitable types of property which can be used to help us perform categorisation, the most obvious of which are likely to be perceptual categories, particularly those based on sight, such as the shape and size of an object, as well as its smell and any noises which are associated with it. We shall see in chapter 3 that there is experimental evidence (Landau, Smith, & Jones, 1988) to show that children use certain perceptual properties, such as shape, as opposed to others like texture.

When encountering an object, we compare it to categories which have already been created, and decide whether or not the new object is a member of any of the existing categories, or whether a new category must be created to account for the object. In this way, categorisation is at its most basic a division of things into two sets: things which are X and things which are not X. This dichotomy between X and X', which underlies the important semantic principle of antonymy as described above (Lyons, 1977), is the basis on which many models of categorisation and meaning have been built, with a clear distinctive boundary between members of the category and non-members. In this section we will explore different theories about the nature of categories, and what they are made up of.

2.3.1 Classical Categorisation

Aristotle (350 B.C./1933) made the distinction between the essence of a thing, made of the properties which define the category, and its accidents, which are incidental properties not used in categorisation. The essential properties make up, in this classical view, a set of necessary and sufficient features to define the category, which makes a clear decision on category membership. For instance, a square is a closed two-dimensional shape of four sides which are equal in length and arranged at right-angles, and the property SQUARENESS is necessarily and sufficiently described by these features. If an object does not have four sides, is not closed and two-dimensional, the sides are not equal in length or not arranged at right-angles, then it is *necessarily not* a square. If an object does have four sides which are equal in length and arranged at right-angles, and is two-dimensional and closed, then it necessarily *is* a square. Other features, such as the colour of the object, what it smells like, whether it appears and disappears from view, are unimportant in deciding whether it is a square; the above properties are *sufficient* for use to make an accurate decision.

Classical categorisation appears straightforward, elegant and plausible, and is very good at providing definitions in particular fields, such as mathematics (as we have already seen), family relations (your *sister* is a female who has the same parents as you), and law (*theft* is the dishonest appropriation of property belonging to another with the intention of depriving that person permanently of it). All these fields are to some degree technical fields, in which strict definitions are extremely important, although arguably not always agreed upon, particularly in the case of legal categories, and these concepts which are well served by classical categorisation are known as *nominal kind concepts* (Schwartz, 1980).

In contrast, classical categorisation turns out to be difficult to use in the definition of *natural kinds*. For instance, all English speakers understand the category DOG, and understand that dogs are different from cats, horses and pigs. However, it is notoriously difficult to come up with a classical *definition* of a dog. If we observe dogs, and come up with a description based on their perceptible features, we could start with shape, colour and texture. But given the massive variation even in terms of these features of objects which we would categorise as dogs, it is not clear to see how any categorisation based on the intersection of these features could clearly divide things into dogs and non-dogs, nor is it easy to see which other features should be added to the list in order for the categorisation to be any more successful. On the other hand, natural kinds such as dogs are clearly defined to some extent by their perceptual features, and also by more abstract features which may correspond to their actions and behaviour, for instance barking and chasing their tails. We have run into what Wittgenstein (1953) called the *vanishing intersections* problem. Wittgenstein failed in his attempt to created a feature-based definition for GAME, and concluded that there was no combination of features which all games had in common. Instead, however, he proposed that different examples of games were actually related by having what he called 'family resemblances' to each other.

2.3.2 Prototype Categorisation

Other investigators have since strengthened Wittgenstein's findings about category membership on the grounds of family resemblance. Rosch (1973) famously showed that category membership is in fact graded or fuzzy rather than all-or-nothing, and also that people are happy to think of some things as *better* examples of a category than others. For instance, subjects were quicker to classify a robin as a bird than a duck, and a table more readily than a stool as a piece of furniture.

Further investigations by Labov (1973) showed that subjects did indeed make use of features to define categories on man-made artefacts, but that these were not necessarily just perceptual physical features inherent in the objects, rather they could also be derived by the language user, such as the presumed purpose of the object[4], and projected back onto the object. For example, subjects were more likely to categorise an object as a cup if it contained coffee, but would categorise the same object as a bowl if it contained mashed potatoes. Importantly, no one attribute is decisive in terms of category membership, but its presence simply increases the probability that the object will be categorised in a particular way. We can see this also with natural kinds, where even a central attribute of birds, namely the ability to fly, can be overridden by the presence of enough other features, so that ostriches and penguins are still included as members of the category BIRD.

Under this theory, a *prototype* refers to the best representative member of the category, and the category membership function is more reminiscent of a probabilistic view or a weighted sum of properties rather than the straightforward combination of binary features in the classical view. Another way of looking at the prototypical view of categorisation is in terms of *exemplars* corresponding to particular, prototypical, instances of the category, against which prospective members are measured for similarity, again according to some function which weights the relevant properties. Prototype categorisation seems to work well not only for natural kind concepts, but also for artefact concepts such as furniture, described by Rosch and Mervis (1975), and also moral concepts such as 'kindness'.

Promising as the prototype view of categorisation is, it also has its own difficulties. For instance, we appear to make use of prototypes as a model for categorisation even when the categories are clear and distinct. In addition, an important feature of meaning is the property of *compositionality*, where the meaning of the whole is made up of some function of the meanings of the parts. The existence of compositionality is one of the distinguishing

[4]Remember that these objects are all artefacts, and so it is reasonable to assume that there is some purpose behind their creation.

features of human language, which explains both its productivity and its systematicity, and yet prototype concepts can't apparently easily be composed into complex concepts.

Inappropriate Prototypes

Some nominal kind categories, as we have seen, do have a straightforward classical description in terms of necessary and sufficient features, particular geometric and other scientific categories such as *triangle* or *parallelogram*, yet equally we can say without any difficulty that some triangles are more prototypical than others.

Most interestingly, Armstrong et al. (1983) used Rosch (1973)'s methods of judging category membership to investigate odd and even numbers. Bizarrely, given the objective scientific definitions of 'prime number', and of the even more familiar 'odd number' and 'even number', together with the distinct and well-understood membership criteria for these categories, Armstrong et al. found, for instance, that 3 is rated a better odd number than 247, and a better prime number than 2. Even concepts which can be easily defined by necessary and sufficient features, therefore, also have prototype structure which would seem to be redundant.

Similarly, fuzzy membership of categories is problematic. Although we can understand why robins are considered 'better' birds than ducks or ostriches, and readily agree that robins are more central to the category, nobody would actually agree that ostriches are not birds. There is still a distinct, classical-like boundary between birds and non-birds, and there is no debate about which side of this boundary ostriches fall, whereas a true fuzzy or vague boundary would surely lead to differences of opinion about category membership.

Non-compositional Prototypes

Fodor (1998a) attacks the whole prototype-based philosophy of meaning because of their non-compositionality. One of the most interesting properties of human language is *compositionality*, the fact that each utterance in the language is composed of parts, and the meaning of the whole utterance is given by the meanings of the parts together with the way in which the parts are put together. It would seem that any theory of meaning must be able to account for the compositionality of language, yet Fodor claims that no prototypes exist for composite concepts, and demonstrates this with two problems in particular, which he terms the *uncat problem* and the *pet fish problem* respectively.

Firstly, he considers the complex concept NOT A CAT, which is both composed of primitive concepts, each of which has prototypes, and also has satisfiable membership conditions. Despite this, Fodor argues that there is no prototype for the concept NOT A CAT, nothing which UNCATS have in common except that they do not fall into the category CAT. He goes on to say that UNCAT can only be defined logically, as part of some theory of categorisation which includes logical operators and operands, and the means to put them together. There are of course an infinite number of such logical complex concepts, made up not only of negations, but also disjunctions TREE OR CAT and implications CAT IF TREE. It is worth noting, however, that our sketched model of meaning based on binary opposition, on dividing the world up into two complementary meanings, has logical negation built into it, so that the definition of any word also provides a straightforward way to get to the definition of its complement.

There is something artificial and frankly weird about the proposed logical concepts NOT A CAT, CAT IF TREE and others, which might lead us not to expect that there should be prototypes of them. Having said that, it may not be strictly true that there is no prototypical UNCAT, as illustrated by the following exchange from the television programme Black Adder the Third, when Edmund and Baldrick are attempting to re-write Samuel Johnson's dictionary, which they have accidentally burnt (Curtis & Elton, 1987):

> Edmund: "And your definition of 'dog' is . . . ?"
>
> Baldrick: "Not a cat."

Baldrick's definition is of course not very accurate, but it does capture one crucial part of what being a dog is about. The set of features making up a prototypical dog does seem to include something like 'not a cat', and this accounts for a large part of the humour in the above exchange. Certainly in terms of word association, if people are asked to provide an example of a typical something which is 'not a cat', it is very plausible that a large number will answer 'dog'. It seems likely, however, that 'uncat', and conversely 'undog' are exceptional examples, bound up in their close relationship to human experiences. Cats and dogs share many syntagmatically normal patterns (they occur in the same position in the same kind of sentences) and this leads to them having a greater semantic affinity than between two randomly-chosen nouns (Cruse, 1986). In addition, of course, they constitute a commonly formed pair, not only in semantic contexts where they both appear as prototypical pets, but even in fossilised lexical phrases, such as 'it's raining cats and dogs'. This aside, Fodor's point still stands in general, as it is unlikely that there would be much of a consensus on what constitutes a prototypical 'unladder'.

The second problem raised by Fodor (1998a) is the concept of a 'pet fish', which is, on the face of it, composed straightforwardly out of the intersection of two concepts, or the overlap between being a pet and being a fish. In terms of prototypes, however, this compositionality also appears to break down; although we may agree that cats and dogs are prototypical pets, and trout or salmon are prototypical fish, there is no obvious way to convert these prototypes into the prototypical pet fish, which would likely be something like a goldfish. Furthermore, the features of this complex prototype doesn't seem to be constructed from the features of the simple prototypes. The features of a prototypical pet fish might be that it is small, brightly-coloured, lives in a transparent bowl and the like. None of these features would appear anywhere in a list of features describing either a prototype pet, or a prototype fish.

2.3.3 Naïve Essentialism and the Theory Theory

So both the classical view of necessary and sufficient features, and the prototype view of fuzzy membership and measure of typicality are elegant in some respects and problematic in others. Both kinds of categorisation seem to be used in different circumstances, as some categories have a straightforward classical definition, while others fall foul of the vanishing intersections difficulty and can only be defined in terms of prototypes.

Keil (1992) tries to square this circle by saying that there are different kinds of concepts, and the way in which concepts are represented changes as we develop through childhood. The two broad classes of things are nominal kinds, which have defining features, as in the classical categorisation system, and natural kinds, which are more inclined to have characteristic features, as in the prototype theory. Crucially, for Keil, concepts are embedded in *theories*, leading to his ideas being labelled the 'theory theory'. Bloom (2000) takes this idea further, by suggesting that categorisation originates from a child's naïve reasoning about the world, which is bootstrapped by feature similarity measures and the correlation of particular features in the world, but then expanded on through learning from interaction with the world, as the child builds explanatory models, or theories, of how the world works. This *naïve essentialism* occurs as we notice the similarities of properties and actions on objects, and try to rationalise them, concluding that there must be an underlying, imperceptible property, a feature which does explicitly define membership of the category. The theory theory differs from prototype theory because the features used are not just those which are perceptible, and categorisation is not based on some weighted measure of similarity, but instead the features include *essence relationships* which explain occurrences and correlations in the world.

There appears to be some potential circularity involved in defining a category in terms of an imperceptible property which exists only to define the category, and it begs the questions of which theories (from an infinite set of possible theories about the world) are actually entertained by the categorising child, why only those theories, and how decisions between the theories are made. On the other hand, appealing to a property which defines the *essence* of the object does account for the shift in conceptual development from using characteristic features based on perception to using defining features based on essence, known as the *characteristic-to-defining (C-D) shift*.

Keil (1992), for instance, experiments with children using cases where objects of natural kinds are outwardly transformed, but essentially (in their essence) unchanged. He presents an imaginary raccoon which has been dyed black, with a white stripe down its back, and implanted with a sac of smelly odour, so that it looks and behaves like a skunk, even a prototypical skunk. Interestingly, younger children categorise the animal as a skunk, relying on the characteristic features with which they have been presented. Older children, on the other hand, insist that the animal is still a raccoon, relying on 'essential' features to make a decision, even though the children are only presented with characteristic features. On further questioning about why the animal is a raccoon despite all that has happened to it, the older children insist that, because it is alive, its essence cannot be changed.

Furthermore, Keil (1992) found that this dramatic developmental shift in children is gradual and occurs at different times for different types of concepts, even into adulthood, as opposed to Vygotsky (1934/1986)'s earlier assertion that there was a universal developmental shift across all types of concepts, from the child's use of exemplars to the adult's use of definitions. It is also claimed that this kind of C-D shift is used more generally, in that we are likely to use characteristic features to make an educated guess at categorising an object, though if forced to make a final decision we would insist on waiting until the defining features were known.

It appears also to be true that as we collect information about the world and investigate more properties of the objects in the world, we can and do revise our decisions about category membership. This occurs not just on the timescale of an individual's lifetime, but over longer timescales, as the collective knowledge of a community grows and changes. For instance, Quine (1969) gives the example of *fish*, which would have included, until recent scientific discoveries, *whales* and *dolphins* amongst their members. Many such cultural categories recur across different languages, and indeed all language communities structure their thought about plants and animals. Many of these taxonomies are made up of essence-based species-like groupings and rankings of groups, which often do not bear any great resemblance to the current Western scientific taxonomies (Atran, 1998), yet others are classified according to how the animals and plants fit into the lives of the people. In the Kalam language spoken in Papua New Guinea, for example, animals are divided into many groups, including *kopyak*, or rats and mice found near homesteads which are considered dirty and disease-carrying; *as*, including frogs and some small marsupials and rodents; *yakt*, which consists of flying birds and bats, and *kmn*, a generic word for all marsupials and rodents, excluding dogs, pigs and mammals in the other categories mentioned (Pawley, 2001).

On the other hand, Western scientific taxonomies are often ignored by the general populace, even if they are scientifically accepted and uncontroversial. For instance, Dupré (1983) claims that there is no scientific distinction between onions, garlic and lilies, as they are all part of the same family *liliaceae*[5]. Despite this botanical similarity, Dupré also is quite correct in emphasising, however, that there is a substantial difference in most Western cultures between sending someone a gift of white flowers and a set of onions. Even if there is no scientific basis for a distinction to be made between lilies and onions, or between small marsupials and larger marsupials, we can find numerous examples where humans create essence-based folk distinctions between groupings which are culturally important.

Keil (1992) concludes that the creation, modification and entrenchment of theory-based categorisation is used even by very small children, who despite the C-D shift, do not categorise *solely* on the basis of perceptible features. Ontological judgements about what *kind* of object a thing is become the basis for the theories we create to explain what features might be necessary to distinguish objects of that kind.

2.4 A Web of Meaning

As we have already briefly discussed, if asked what a word 'means', we naturally turn to other words, with which we try to paraphrase our target word, to make it clearer by the use of more intelligible words. Even this straightforward task, the clarification of words' meanings through other words, however, is much more difficult than might initially be imagined. The following definition of 'sneeze' is taken from the Oxford English Dictionary (OED Online http://oed.com/cgi/entry/00229097) (Simpson & Weiner, 1989):

[5]Actually, it is rather strange to say that there is no scientific distinction; although in the same family, if we go down a level of classification, onions and garlic on the one hand are in the genus *allium*, while lilies are in the genus *lilium*. Dupré's point that non-botanists do not think of them as related still stands, however.

> *sneeze, v.* To drive or emit air or breath suddenly through the nose by an involuntary and convulsive or spasmodic action, accompanied by a characteristic sound.

I suggest that it would be quite surprising if very many speakers of English would be any the wiser if they were reading this definition in a state of ignorance, trying to understand what 'sneeze' meant. Many of the words used in the definition (e.g. convulsive, spasmodic, involuntary, emit) are much less familiar and therefore probably less intelligible than 'sneeze' itself. On looking up 'convulsive', we find it defined only as 'characterized by convulsion', which is itself defined as follows (OED Online http://oed.com/cgi/entry/00049329) (Simpson & Weiner, 1989):

> *convulsion, n.* (usually plural) an affectation marked by involuntary contractions or spasms of the muscles . . . ,

which refers back in a circular fashion to derivatives of two of the less familiar words (involuntary, spasm) in the original definition of 'sneeze'. Not only are the words used more confusing than clarificatory, but certain parts of the definition rely on the user understanding the target word in the first place in order to make any sense at all. The ignorant reader of the first definition, for example, is told only that a sneeze is accompanied by 'a characteristic sound', but is given no details at all as to the nature of this sound, or how to recognise it.

Goddard (1998) gives many similar examples of circular definitions found in all monolingual dictionaries, where one word is explained in terms of another word, which is in turn explained in terms of the first word. Sometimes these chains are longer than two words, but in the end they are inevitable, if we assume, as is generally the case, that the purpose of dictionaries is to try to describe every word in a language. It should be clear that this is an impossible task without circularity.

An alternative to circularity would be to relax the aim of describing every word of a language in a dictionary, and instead to leave a core of undefined fundamental words, whose meaning is assumed and which can then form the building blocks for other definitions in the dictionaries. Of course, we would like the number of words in this undefined core to be as small as possible, yet expressive enough that all other words can be defined in terms of only these words and others defined from them. Wierzbicka (1996) has developed to this end a universal natural semantic metalanguage from a universal list of semantic primes, the exact composition of which is still under debate and revision, but certainly including primes such as *you, the same, good, below, because, a long time*. It is claimed that the simplest sense of all these words or phrases are lexically universal and can therefore be matched across every human language by either a word, a phrase or a bound morpheme.

However, many people are sceptical of attempts like Wierzbicka's to find universal semantic primes, and we shall see later, in chapter 3, that in certain well-defined semantic domains, such as the organisation of spatial relationships, which has been studied cross-linguistically at great depth, it is very difficult to find any common ground between all languages, even those which are closely related to each other.

2.4.1 Entering the Web

Meanings, therefore, exist in a complicated web, where each meaning is related to, and connected to, many others. The connections between the nodes on this web correspond to different kinds of semantic relationship. As previously mentioned, the fundamental problem in trying to work out what utterances mean is that the word forms themselves don't appear to have an intrinsic meaning and exist in a different medium to that of meaning. If we assume that a word token ('mortgage', for example) is inherently meaningless, how can rephrasing it in terms of other meaningless symbols ('loan','interest', etc.) explain the meaning to us? We are just moving around our web of meaning in circles, and although we can do this effortlessly, as popular word association games show, it does not help us to work out how we get into the web in the first place, so that we can begin to forge the links between words and meanings. It seems that there are two obvious possibilities:

1. either (some of) the web of meaning is innate and already in our brains, so we don't need to get into it;
2. or (some of) the meanings are grounded somewhere in the real world.

2.4.2 Innate Meanings in Animals

The predictability of the environment in which an animal has evolved, and the representation of that environment which it has found useful over an evolutionary timescale would seem to provide an initial starting point to investigate whether it is plausible that meanings could be innate. After all, every animate creature has to have some kind of

Predator	Sound	Response
Snake	Chutter	cluster together and stare into grass
Eagle	Cough	run into bushes for cover
Leopard	Bark/Chirp	run up into trees to ends of branches

Table 2.1: Vervet Monkey Calls and Responses (after Cheney and Seyfarth).

model of the world and things they encounter in it, whether they are born with this model or build it themselves.

Bickerton (1990) reports experiments in which some monkeys, raised in isolation, show signs of alarm when they first come into the vicinity of anything which bears some resemblance to a snake, strongly suggesting that they have an innate representation of snakes. It is clear why such an adaption would be useful from an evolutionary point of view: other things being equal, baby monkeys who did not need to learn to avoid snakes would be more likely to survive and reproduce their genes than those who had to discover the danger from experience and ran the severe risk of being killed. Could an innate representation system such as clearly exists in these monkeys be the basis for the conceptual system on which human language is built?

One of the most famous studies of non-human communication systems was made by Cheney and Seyfarth (1990) in their study of vervet monkeys referred to by Bickerton above. Cheney and Seyfarth discovered that the monkeys make different calls in response to different predators. The vervets have three main groups of predators: snakes, eagles and leopards and other cats. On seeing a predator, the monkeys emit a particular cry, which alerts the rest of the group. Crucially, the vervets' escape strategy needs to be different for each predator, and they do indeed react differently to each signal to evade the predator, as shown in table 2.1. The monkeys clearly have some kind of representation of each of their predators, and they also seem to use arbitrary calls which stand for these representations. The alarm calls could be called signals, which *mean* 'snake', 'eagle' and 'leopard'. Alternatively, we might like to give the calls less noun-like meanings, and instead say that they mean holistic expressions such as 'cluster together and look in the grass!', 'run for a bush!' and 'run for the end of a branch!'.

Furthermore, Cheney and Seyfarth claim that while the representation is innate, it is also tuned by experience. Young vervet monkeys make the same calls as the adults, although initially they bark whenever anything is approaching on foot, and cough when large leaves fall from the sky. Only later do they narrow down the range of 'meanings' which are appropriate to each call. From a more sceptical point of view, Budiansky (1998) points out that youngsters making errors will stand out to observers, while it is difficult to observe them making the correct call, simply by sheer weight of numbers; moreover, when a real predator arises, it is more likely that it will be an adult who first notices the intrusion and sounds the alarm.

Similar phenomena have been noted in other animals, who also make different calls to respond to different situations. Morton and Page (1992) describe how ground squirrels have two main predators: birds of prey like hawks and mammals like badgers, which rely on stealth to catch the squirrels. In the same way as the vervets, ground squirrels need different strategies for avoiding their predators: for badgers, they need only to stand tall, displaying that the predator has been spotted, but for hawks they need to dash for cover as quickly as possible. As we might expect, the squirrels also have different calls to respond to each of these predators, and these calls have been therefore been interpreted as *meaning* 'hawk' and 'badger'.

Indeed, animal calls have often been taken as proof of representational capacity (Cheney & Seyfarth, 1990; Schoenemann, 1999), though this is not uncontroversial. There are, in fact, qualitative differences between the animal systems and human systems. An important difference is that animals do not use the calls without the predators being present (Burling, 1993), nor without an audience of other conspecifics (Cheney & Seyfarth, 1990). The animals' use of signals, therefore, could be profitably compared to humans using the signal of a sneeze to symbolise pepper. Whenever our noses come in close contact to pepper, we sneeze, alerting others that there is pepper in the vicinity. But sneezing is involuntary and cannot be controlled, so we cannot refer to pepper by sneezing when it is not there. Likewise, the vervet and ground squirrel calls are a reflexive response to certain conditions in the environment. Budiansky (1998) points out that the ground squirrel's 'hawk' call can just as easily be interpreted as an evolved mechanism for confusing the hawk; the calls interfere with the attacking hawk's locating of its prey. Hawks are generally very good at locating the source of a call, pinpointing it almost instantaneously, but the appropriate ground squirrel calls actually make the hawk look 90 degrees in the wrong direction, the hawk's temporary confusion giving the squirrel vital seconds to make its escape.

Furthermore, Budiansky makes the important point that giving an alarm call when you are alone just draws attention to yourself, setting yourself up to be killed, but in the company of others you can use the call to recruit others to repel the predator by mobbing (as the vervets do with snakes), or to create pandemonium by all running for cover at once, thus making your own escape less conspicuous (as the vervets do with both eagles and leopards, but to different destinations). The fact that vervet monkey calls are only used in these circumstances, in the presence of both the predator and an audience, strongly suggests that the vervets' concept of 'eagle' is only brought to mind, as it were, by sensory recognition, rather than by any process which we might term *thinking* about eagles.

To conclude, then, vervet monkeys, ground squirrels, and other animals do indeed appear to have evolved a strategy for predator avoidance which involves alerting other vervet monkeys, but it isn't a system for communicating ideas, in contrast to human language. One vervet monkey is not trying to communicate the idea of EAGLE to another one, like English speakers do when they use the word *eagle* or Hungarian speakers with the word *sas*, but instead the first monkey is automatically reacting to the sight of the predator, and the second monkey is, in turn, automatically reacting to the sound the first monkey makes. On the other hand, vervet monkeys certainly *do* seem to have an innate representational system. Given the evolutionary development of humans and monkeys from a common ancestor, and the massive sharing of genetical material, could some kind of innate representational system be present in humans?

2.5 Human Concept Acquisition

There is, however, an obvious difficulty in the size of the human representational system of meaning. We can accept that the vervets calls are innate, and even theorise plausibly about the origin of their representations through basic evolutionary principles of predator avoidance; the human representational capacity, on the other hand, is enormous. Explaining the existence of each of this vast set of human concepts is impossible, yet even if there were a finite limit to our capacity, it would be undesirable to go through each concept in turn, and finding a suitable explanation for its existence; instead, it is the capacity for concept acquisition itself which must be explained.

The origin of meaning has been debated for many centuries, and the crucial dichotomy on which the debate hinges is the problem of concept acquisition, or how we come to know what we know. As we saw in chapter 1 with relation to theories of the nature of language, the two sides of this somewhat rancorous debate are known as *nativism* and *empiricism*. Nativists believe that concepts are somehow genetically-determined, and are acquired by being triggered by experience, while empiricists believe that they are created inductively, on the basis of learning through experience, by forming and testing hypotheses about category membership.

2.5.1 Nativism

One of the most plausible arguments for the innateness of meaning is the apparent intertranslatability of all human languages. For instance, the words *woman*, *donna* and *nő* all have the same meaning, and can be directly replaced with each other when translating between English, Italian and Hungarian. It is arguable that all concepts can be translated into identical concepts in other human languages. And yet, even though this claim appears reasonable, it massively underestimates the richness of human language. Every language encodes untranslatable concepts, which only 'make sense' in the particular cultural setting in which the language was born. We have already seen how Kalam-speakers linguistically classify the fauna in their surroundings, yet can we really come up with a satisfactory English or Swahili translation of a Kalam word like *kmn*, without resorting to an enormous explanatory paraphrase, or, worse, a list of all the animals which are covered by the word and another list of those which are excluded? Goddard and Wierzbicka (1994) demonstrate persuasively, and arguably at odds with their hunt for the universal semantic primes which they believe all meanings can be decomposed into, that this problem is not just one of translation, but also of the ease by which particular thoughts are available in particular languages:

> " ... thoughts related to [Russian] duša, for example, can be formulated in English only with great difficulty and at the cost of cognitive fluency, whereas in Russian they can be formulated more or less effortlessly." (Goddard & Wierzbicka, 1994, p. 59)

We will see in chapter 3, that, at the very least, it is undeniable that the ways in which languages divide up the available semantic space differ greatly, and that many concepts which appear initially to be plausible primes are not used at all by speakers of some languages.

Fodor (1975, 1998a), one of the strongest proponents of nativism, argues against empiricist claims by means of what he amusingly calls the 'Standard Argument'. Firstly, he assumes that most lexical concepts[6] have no internal structure. If concept learning is an empirical process, then in order to create the concept, the learner needs to formulate and modify a hypothesis about category membership. Fodor concedes that this is possible

[6]Fodor (1998a) uses 'lexical concept' to refer to a concept expressed as a word rather than a phrase. This distinction between words and phrases is not particularly useful, especially if it is based solely on data from one language. Wierzbicka (1996) shows how many Australian languages, for example, use a suffix to express a concept which would need a whole phrase such as *for the sake of* in English.

for categories which have easy classical definitions, like BACHELOR, which, he assumes, expresses the same concept as that of UNMARRIED and MALE[7]. But, he argues, it is impossible to learn primitive, unstructured concepts this way. Essentially, it is impossible to form a hypothesis about the meaning of a concept like RED, without using the concept itself in the formation of the hypothesis. This clearly falls into a circular argument, as we must presuppose the availability of a concept in the explanation of how that concept is acquired. Primitive categories, therefore, cannot be learnt and must be explained in some other way. This point is accepted by many, indeed Jackendoff (1990) proposes that we are born with a set of primitive concepts, and the means to combine these concepts to form an infinite set of complex concepts. Fodor's main point, however, is not just that primitive concepts are unlearnable, but that *most* concepts are unstructed, primitive, and therefore unlearnable.

Radical Concept Nativism

This theory of massive innateness, or *radical concept nativism* is controversial, to say the least, as it requires that *all* primitive concepts are innate. Given that Fodor (1998a)'s definition of a primitive concept is so broad and all encompassing, we are left with the frankly bizarre conclusion that even concepts like MODEM and QUARK are innate. Even worse, we can of course invent new names for objects and actions at will, and yet Fodor appears to believe that there is some limit to this ability, that at some point our potential stock of names will be exhausted. If this has not stretched plausibility to breaking point, then we find ourselves, from an evolutionary perspective, back in a situation where we need to explain the existence of each and every concept in turn, unless we can find a reasonable story for the general acquisition of concepts. Although Fodor (1998b) is critical of evolutionary explanations for these kind of phenomena, he is well aware of the general unpalatability of radical concept nativism, and so develops a concept acquisition method which is not based on learning through rational hypothesis generation and testing, but is instead 'brute-causal', and based on what Fodor terms *triggering*.

Although a nativist account might seek to downplay the role of experience in concept acqusition, this is not how Fodor (1998a) views things. Indeed, in his earlier work on the language of thought (Fodor, 1975), he emphasises that expressions have extensional semantic properties, that they denote objects in the world. Fodor (1998a) tries to square this circle by appealing to the 'triggering' of innate concepts. This notion is as simple as it sounds; certain specific inputs trigger the availability of certain (pre-defined) concepts.

[7]Of course, BACHELOR and UNMARRIED MALE do not denote exactly the same category, as can be seen if we try to decide whether a baby boy, a tom-cat or the Pope is a bachelor. But Fodor's argument does not rest on this point.

Under this mechanism of acquisition, the whole process is emphatically not rational, but is 'brute-causal'. Acquiring a concept is the mind becoming locked to the property which that concept expresses.

Fodor's idea of triggering appears to be reminiscent of the phenomenon of *imprinting* in very young animals. Lorenz (1966) discovered how young ducklings and goslings learn to follow their parents soon after they are hatched. The young birds respond to visual and auditory cues from their parents, and these cues trigger a brute-causal response which affects the behaviour of the young birds for the rest of their lives. Lorenz discovered, in fact, that the birds respond to the first conspicuous moving object they are exposed to; Lorenz himself imitated the call of a mother goose in front of newly hatched goslings, whereupon they followed him around as if he was their mother. Imprinting occurs only in a very short *critical period* soon after hatching; once the duckling has identified the features of its mother through imprinting, it discriminates all other objects from the familiar one, which causes it to shrink away from the other objects and towards the familiar. Young birds are imprinted not only with the characteristics of their mother (filial imprinting), but also with those of their siblings (sexual imprinting), which influences their mate preferences when they are adults. Ducks who are imprinted on human experimenters, therefore, will try in adulthood to mate with humans[8]. It certainly appears, therefore, that ducklings have at least one innate concept, which we could anthropomorphically describe as MOTHER, and that this concept is triggered by the ducklings' experience of a certain input, through a process which is certainly not rational. This seems to be a similar scenario to that suggested by Fodor (1998a) for human acquisition of concepts, although on a much larger scale; rather than the triggering of one or two concepts through particular experiences, we must remember that Fodor argues that most of the enormous number of human concepts are acquired in this way.

So how does experience of a particular stimulus cause the *human* mind to become locked to the particular property? Under the brute-causal approach, experience causes the acquisition of concepts, but this is not based on confirmation or denial of any semantic hypotheses, and it is therefore arbitrary. The problem for a truly brute-causal, non-rational approach is that there is an infinite set of causal relations which could therefore theoretically trigger concept acquisition. Only certain of these causal relations, however, do apparently lead to concept acquisition; worse still, those relations which do lead to concept acquisition do actually appear to be extractable from the environment on some kind of rational, empirical basis.

[8]In a non-experimental scenario, of course, ducks who imprint on things which are not members of their species will be extremely unlikely to survive anywhere near long enough to reach sexual maturity.

This problem manifests itself famously as the doorknob/DOORKNOB (d/D) problem: why is it experience with doorknobs which lead to the acquisition of the concept DOORKNOB, rather than experience with giraffes or whipped cream (Fodor, 1998a)? The rather unconvincing answer to the d/D problem, according to Fodor, is that doorknobs end up being things which have the property which cause human minds to acquire the category DOORKNOB.

> "*[D]oorknobhood* is the property that one gets locked to when experience with typical doorknobs causes the locking and does so *in virtue of the properties they have qua typical doorknobs*" (Fodor, 1998a, p.137, emphasis in original).

Typical doorknobs, of course, characteristically cause the acquisition of the concept DOORKNOB, leaving the argument, despite his protestations to the contrary, and despite his logical manœuvres through (proto-)typical doorknobs, decidedly circular and uninformative. The sheer scale of the number of trigger receptors which must be, according to Fodor's account, waiting for the appropriate trigger in order that they make a concept available, make this kind of brute-causal acquisition unrealistic.

A further problem to Fodor's brute-causal acquisition of concepts is that many concepts are not actually acquired through experiences of the 'things-which-cause-concept-locking'. Instead, concept acquisition is mediated in many cases, very probably most cases, through language. We cannot seriously entertain the suggestion that the concept QUARK is acquired through experience with quarks. Many people have no concept QUARK, but this is not because they haven't experienced quarks, rather that they have not had the information, via a book or lecture or via language of some other sort, of what quarks are[9]. On the other hand, many people *do* have the concept GOD, and although most would argue that their having the concept has indeed come about through experiences which may have triggered the availability of the GOD concept, others would deny that Fodorean triggering is possible with this kind of concept, yet would allow that people do possess the concept[10].

Putnam (1975) shows interestingly how there are also many concepts which we have some knowledge of, but nothing like enough to actually use. He gives the example of ELM and BEECH; although he knows that both refer to different types of tree, he does not

[9] Arguably, most people who *do* have the concept QUARK probably have no more defined a concept than 'some kind of (sub-atomic) particle'.

[10] Indeed, atheists, although denying the existence of gods, must have the concept GOD in order to deny that it exists.

know how to discriminate one from the other. Even given two trees, an elm and a beech, he could not (and neither could many people) indicate which was which, with anything greater than chance accuracy. His proposal for how such a system works is through a division of labour, whereby people use terms without knowing their full meaning, while acknowledging that the meaning does exist somewhere in the language community.

Boyer (2000) has investigated the evolution and creation of religious concepts in cultures throughout the world in terms of their connection to inuitive ontology. In particular, he shows how very common religious concepts such as 'there is an omnipotent person who knows everything we do' include explicit violations of how the world works, as well as activating a background of default expectations which are not violated. For example, 'there is an omnipotent person who knows everthing we do, but then forgets it immediately' is not an appropriate religious concept, though it is of course supernatural and not obviously absurd. Cultural concepts such as these, which are very prevalent in human society, cannot be acquired through any kind of triggering experiences, but must instead be acquired by some other mechanism of connection with the concept's referent. It seems plausible that language is the most obvious candidate for the mechanism which has made possible this phenomenon, which has been called *reference borrowing* (Devitt, 1981), and *symbolic theft* (Cangelosi & Harnad, 2000). Importantly, it is distinctly rational or psychological in origin, therefore posing a serious problem for the brute-causal mechanism of concept acquisition espoused by Fodor (1998a) which rejects any form of rationalism in concept development.

2.5.2 Empiricism

In contrast to nativism, empiricists claim that concept acquisition is based on interaction between the learner and the environment, and that there is no feedback from any other human, who might be regarded as a teacher. If there is any sharing of categorisations, then these are due to both the cognitive architecture of the brain, and the biases present in the environment, both in the structure of the world and in the particular exposures of a learner. We shall see in chapter 8 how these environmental effects can indeed have a large impact on both sharing of meanings and on the success of communication systems.

Instead of concept acquisition by triggering, which essentially maps certain experiences, or generalisations of experiences, to innate concepts, empiricists would contest that concept acquisition is more accurately represented as *concept creation*. Introducing this notion immediately forces us to confront two important questions: how are concepts created; and which concepts are created? Fortunately, there are straightforward answers to both questions: concepts are created to allow an individual to make sense of the world, to recognise and discriminate situations from each other; and the particular concepts which are created are those which are *useful* in some way to those who create them, because they allow the discrimination of situations and objects. We can appeal to the development of these capabilities on straightforward evolutionary grounds; the ability to distinguish predator from inanimate object is clearly of utmost importance, and a creature unable to do this will not survive to pass on its genes.

Human concepts are, of course, characterised by their variety and flexibility, which poses difficulties for a nativist account of their origin. If we imagine a group of human babies living on Mars (assuming they could survive), then we could all imagine how they would, without doubt, develop words for Martian situations and objects, and how they wouldn't have need for concepts like TREE or WATER, which refer to things which don't exist on Mars. We can see that if meanings are formed through triggering in the environment, the children would never obtain meanings for TREE and WATER, but we return again to the problem of the trigger receptors which must be ready to make Martian meanings available, just in case the children had happened to be born on Mars.

As we have seen, however, concepts are not just created in response to our environment, but also as a result of communication and interaction between humans, mediated through language. Cangelosi and Harnad (2000) present a model world where agents interact with mushrooms to discover whether they are poisonous, where agents who are allowed to 'steal' symbolic categories through reference borrowing, substantially outperform agents who have to learn the hard way through experience alone. Although the initial categories are acquired by trial-and-error, and are grounded in the world, once the categories are *named*, the authors suggest a situation where the new categories can be swapped by conversation, allowing both parties to increase their knowledge, and their view of the world, much faster than by trial-and-error alone. Cangelosi and Harnad show that symbolic theft has a selective advantage, although interestingly they point out too that symbolism alone is unstable — if there are no toilers in the population to get the symbols from, then the symbolisers cannot get any knowledge and their flexibility is in vain.

These results would suggest that a system of symbolic knowledge, which is initially grounded in reality, is the best of both worlds. Language users can boot their systems by acquiring them with reference to the world, and then new categories can be built from the grounded categories, assuming we allow for the composition of categories. The crucial insight which enables the development of languages is that of *symbolisation* (Deacon, 1997), although how symbolisation first occurred remains a matter of much conjecture.

Cowie (1999), one of the foremost empiricist philosophers of meaning, gives ground to nativism by accepting that some concepts might be innate, but is insistent that most 'higher-order' concepts are learnt. She argues persuasively that there are highly sophisticated psychological processes going on by which we grasp the meaning of some concepts like XYLOPHONE and PLATYPUS, which Fodor (1998a) claims are triggered. We learn from our experience, and use existing, innate, concepts in order to construct definitions and prototypes, and to learn how to use reference transfer. These kinds of intentional mechanism are necessary, as we have seen, to account for the acquisition of concepts such as QUARK and GOD, which it is implausible to try to account for in terms of brute-causal triggering by the environment.

Definitions and prototypes, therefore, together with reference transfer, once created from basic concepts, serve to allow the construction and learning of further concepts, and so the process of concept development continues indefinitely. To some extent, Cowie and Fodor, despite their withering criticism of each other, are arguing for the same position, i.e. there are some basic concepts which are innate, and other, more complex concepts which are learnt. The difference is one of degree: Fodor claims that nearly all concepts are primitive, while Cowie claims that most concepts are structured.

Cowie (1999) makes a distinction between *intuitive meaning* (what you need to know in order to have a concept) and *technical meaning* (semantic properties which fix the concept's reference in the world), in order to overcome Fodor (1998a)'s objection about the non-compositionality of prototype meanings. Having made this distinction, she argues that prototypes are the intuitive part of meaning, they are the things you need to know to have the concept, but they do not fix the concept's reference. They are non-compositional, as Fodor (1998a) argues, but this compositionality does not apply to the intuitive part of meaning. Although it is plausible that in order to have the intuitive concept of BIRD, you need to have a prototype which is built up of more basic concepts (namely is feathered, can fly, etc.), it is not true that this prototype fixes the reference of BIRD. As we have already seen, there are feathered things which can fly which are not birds, and vice versa. Crucially, although the prototype of BIRD is made up of other concepts, it does not presuppose the existence of the concept BIRD, and so it can be learnt.

Again, we discover that classical and prototype approaches to meaning both have their place, and so it seems reasonable to adopt Cowie (1999)'s argument that concepts are made up of both an intuitive part, which is based on prototypes, and a technical, reference-fixing part, which is based on classical definitions.

2.6 Summary

In this chapter, I have looked at the nature of meaning, the web in which meanings exist and the relationships which hold this web together. I have looked at different models of categorisation, and conclude that despite the difficulties in explicitly defining category membership in terms of necessary and sufficient features, by proximity to a prototype, or on the basis of some imperceptible essence, categorisation itself is as solid a rock as we can find on which to build a model of meaning creation. I have explored the nature of concept acquisition, looking at the competing claims of nativism and empiricism. In principle, it is possible to have some innate concepts, (indeed, even the most ardent empiricist (Cowie, 1999) would agree that *some* concepts are innate) if the things they referred to were guaranteed to be of use for the vast majority of people. These meanings might not necessarily refer to particular things like TREE and WATER, but could be more basic ideas of *self*, *objects*, *events*, *individuals* and *kinds*.

On the other hand, I would argue that the majority of meanings are neither innate, nor triggered in a Fodorean brute-causal manner by the environment, but are instead created in response to experience. It is clear that the precise details of our human brain are crucial in determining the concepts we acquire, as are the precise experiences that we have in the world. With this in mind, the argument between nativism and empiricism begins to be a little more one of emphasis than of incompatible positions, and it can be characterised broadly as the empiricist's more active meaning acquisition and rational construction being opposed to the nativist's more passive meaning acquistion through brute-causal triggering.

In chapter 3, we will look in more detail at the nature of language acquisition as opposed to concept acquisition. I will investigate how words are attached to concepts so that they can be used in language, and will investigate how these two processes interact with each other. I will, in chapter 5, describe a model of experience-based meaning creation based on the perceptible features of objects and situations. In the end, as Harnad (1990) points out, concepts must be grounded in reality; they must eventually, if we go round the web of meaning long enough, be able to be used to pick out objects and actions which can be pointed at. Because my model of meaning creation and communication will be based on identifying objects and situations in the world, it is natural, while acknowledging that the nature of meaning is more complicated than simple reference-fixing, that it concentrates more on the technical, classical description of meaning, in order to build meanings which agents can use to communicate about things in their world.

CHAPTER 3

Learning What Words Mean

> " 'When *I* use a word,' Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean — neither more nor less.' " (Carroll, 1872/1998, p.190)

3.1 Introduction

This chapter will investigate the problem of language learning, in particular the problem of how meanings become associated with words. Firstly, in section 3.2, I give a gentle introduction to the problems of deciphering unknown symbols and of the paradox of meaning induction in general, then move in section 3.3 onto a discussion of proposals which have been put forward in the developmental psychology and linguistic acquisition literature to account for the fact that learning language actually comes very naturally to children; these proposals include specific constraints on word learning such as the shape and whole-object biases, and the principle of contrast, as well as more general proposals on the socio-pragmatic development of the child such as understanding the intentions of others and the phenomenon of fast mapping. Finally, in section 3.4, I discuss some more linguistic aspects to the learning task itself, showing how difficult it is to provide any sort of unified, universal account of the semantic system underlying all human language, and how a complete account of language acquisition must also explain how the learner manages to induce the particular linguistic system which must be learnt, in addition to actually doing the learning.

3.2 Codebreaking

The creation and development of a conceptual structure which allows us to categorise and distinguish situations and events in our world is only part of the story. As prospective learners of a language-like system, the main problem we face is trying to work out what all these strange sounds mean. Imagine being transported to a foreign country, and being unable to understand the people around you. The sounds they use all appear to sound the same, and moreover they are very difficult for you to pronounce; many of the sounds *you* can make easily don't appear to be used at all in their speech. But fortunately they are friendly people, and keen to communicate with you, so to help you learn their language, they very kindly provide you with a dictionary; the only problem is that the dictionary is monolingual, explaining words in their language in terms of other words in their language.

Everything is unfamiliar to you as a learner of this language, except that the people are speaking a language, and that the writing in the dictionary and the sounds in their speech both represent that language in different ways. How on earth do you break the code and work out what the sounds and signs mean, what systems lie behind their use?

3.2.1 Deciphering Unknown Symbols

The problem described above is very similar to that faced by the first modern scholars of ancient Egyptian, who were trying to decipher the hieroglyphics without a clear idea of what they stood for. After many false starts, an interpretation only became possible after the discovery of the Rosetta Stone, on which a description of King Ptolemy V's coronation was inscribed in two languages, and in three different scripts: in hieroglyphics, in Demotic, and in Greek. Once it was discovered that the Demotic and Greek sections of the stone were translations of each other, it was possible for the French translator Jean-François Champollion to begin to decipher the third script, and work out what the hieroglyphs represented.

Even with this discovery, however, deciphering the hieroglyphics was not straightforward. The researchers were aided by discoveries such as the convention of writing royal and divine names with surrounding ovals, or cartouches. These discoveries provided a starting point for the codebreaking, which then opened up further avenues to pursue. The deciphering would, however, have been impossible without the Rosetta Stone or a similar bilingual document. Robinson (2002) describes the same problem which still exists today under different circumstances; many languages, famously including among others

Etruscan and the language of Easter Island, remain undeciphered, simply because there is no starting point from which the code can be broken.

But let us move back a stage. Even before we find a helpful bilingual text, including a language we do know, to help us start the codebreaking procedure, we need to gain an insight which now appears obvious, but is in truth at the heart of language: the hieroglyphs *stand for something*. Language is a form of *symbolic communication*, in which a signal stands for a concept, to which it is not related in any way, save for the very fact of the linkage through symbolism itself. For instance, there is nothing in the sound of the word *chair* which suggests any aspect of its meaning; nothing in the word *seat* which suggests its meaning, or the fact that their meanings are related to each other. A *chair*, indeed, is known as *cadira*, *kiti* or *szék* by speakers of Catalan, Swahili and Hungarian respectively; none of these words have any relation either to each other or to the meaning of *chair*.

This arbitrary linkage of form and meaning in the symbolic *sign*, as described by Saussure (1916), is at the core of modern linguistics, and is arguably one of the crucial differences between human and non-human communication. The concept of the signifier and signified being joined in a sign is then developed further to produce the notion of *duality of patterning*, where sets of intrinsically meaningless phonetic items are arranged using a system into a mass of complex, meaningful units. Duality of patterning allows tens of thousands of distinct words to be created from a small set of phonemes[1], and this in turn accounts for the enormous expressive power of human language, which is lacking in other semiotic systems.

Importantly, our human minds are so attuned to this kind of symbolism that it is very difficult to envisage a world without symbols, where the only structure in the world which we can discover is related to the co-occurrence and correlation between objects.

3.2.2 Problems of Meaning Induction

The symbolic insight is a crucial insight, but we are still a long way to working out exactly what an unknown symbol stands for. If we look specifically in terms of language, word learning seems like a very straightforward process, one which is, after all, successfully

[1]The size of a particular language's phoneme inventory also varies considerably; there are, for instance, languages with as few as 11 phonemes like the Papuan language Rotokas (Firchow & Firchow, 1969). An upper bound to this range is more difficult to ascertain; the oft-quoted 141 phonemes for the Khoisan language !Xũ or !Kung (Snyman, 1970) has been questioned by Traill (1985), who argues convincingly instead for a cluster analysis of Khoisan consonants, thereby reducing the number considerably. If we accept Traill's analysis, then it is difficult to find more than around 85 phonemes for the recently extinct Caucasian language Ubykh (Catford, 1977).

and effortlessly completed by very small children, and yet we have seen that words and their meaning are only related through some arbitrary symbolic mechanism. The linkage between form and meaning must clearly be learnt and is not genetic, as we know that children grow up learning the particular languages to which they are exposed, not the languages their parents spoke (although of course in many cases these are the same).

The problem of inducing the meaning which a signal is being used to convey was most famously illustrated by Quine (1960), who presented an imaginary anthropologist, who observes a speaker of an unfamiliar language uttering the word *"gavagai"* while pointing to a rabbit. How does the anthropologist know what *"gavagai"* means? On first glance, we might assume that the word means RABBIT, but why do we make this assumption? Quine shows that, logically, this assumption is not correct, and that in fact *"gavagai"* has an infinite number of possible meanings, including ANIMAL, WHITE, RABBITNESS, UNDETACHED RABBIT PARTS or DINNER!

Given all these possible meanings, then, how does a logically-minded anthropologist decide between them? He may look for confirmation of the meaning, perhaps by pointing at other objects, and questioning the native speaker as to whether they, too, are covered by the extension of *"gavagai"*. By collecting more information through further questioning, the anthropologist can reject some hypotheses from those with which he started. But Quine (1960) proves that this will not actually help, because there will always be yet more logical hypotheses which will be consistent with the new set of data; the set of hypotheses which the anthropologist is seeking to reduce is infinite. Quine refers to this as the *indeterminacy of translation*: no matter how much evidence is collated, the meaning of *"gavagai"* will never be determined.

A very similar philosophical problem was described by Goodman (1954), which is known widely as the *grue paradox*. Goodman presents the problem of two people who have both been exposed to a number of emeralds, all of which have been coloured green. One forms the hypothesis that "emeralds are green", while the other the equally logically plausible "emeralds are grue", meaning "all emeralds have been green up until this moment, and they will all be blue hereafter". Of course, the *grue* hypothesis can never be disproved by experience, and so will always be, logically, just as plausible as the *green* hypothesis. We can easily imagine any number of similarly bizarre yet unrefutable hypotheses, and Goodman shows that, under inductive learning, that there is always an infinite set of logical generalisations which can be made, each of which is consistent with the data experienced, no matter how much evidence is accrued.

This problem of meaning induction is exactly that faced by a child acquiring its native language. How does the child know which of the possible meanings are plausible, and reach the correct conclusion from the infinite set of possibilities? In reality, we know that when faced with these kinds of tasks, children react by reducing the number of possible meanings to which they give credence. It is quite possible, indeed, that Quine's set of possible meanings might reduce in practice to RABBIT, without any further evidence being required. But on what grounds does this reduction of possible meanings take place, so that we can overcome the Quinean problem of meaning induction?

This chapter considers the problem of learning how to associate a form with an unrelated meaning, or how to learn the meaning of an unfamiliar word. We also investigate many of the suggested solutions to this problem, and discuss how these fit into our model of the evolution of communication.

3.3 Constraints on Word Learning

In the following sections, I will consider many of the suggested solutions to these learning problems, before looking at how to implement them in our model. Firstly, I will investigate the intuitively attractive proposal that children learn by being taught by their parents, particularly through being corrected when they make mistakes. Then, I will move on to the hypotheses that children have particular biases or predispositions to disregard some theoretically possible meanings, or to prefer some possible word meanings over others. These biases would serve to greatly reduce the set of possible meanings, and crucially thereby make it finite, and therefore the problem itself soluble. In particular, I will focus here on the whole-object bias, the shape bias, the taxonomy bias, the mutual exclusivity assumption, and the principle of contrast. Finally, I will present proposals which appeal to general cognitive principles rather than specific constraints.

3.3.1 Negative Evidence

One common suggestion for how children learn the meaning of words is that they are explicitly taught by parents and teachers. Under this scenario, a child is given *feedback* on its use of words: if it uses a word in the correct manner, it receives positive feedback to encourage further use; if it uses a word incorrectly, it receives negative, or corrective feedback to discourage further use. This kind of learning process is often called *reinforcement learning*, because the learner's actions are reinforced by the feedback from the teacher.

Despite the simplicity of this idea, and its intuitive appeal, the existence of negative evidence is extremely controversial, both in the acquisition of lexical items and of grammar, as can be seen in Morgan and Travis (1989)'s review of psycholinguistic evidence on the matter. Brown and Hanlon (1970) demonstrated that parents did not correct their children when they produced ungrammatical sentences, but did correct them when they produced sentences which were not true, apparently providing some evidence in favour of the existence of corrective feedback at least in some instances. Such occurrences, however, are by no means culturally universal; Lieven (1994), for instance, describes cultures in which parents do not even speak to their children in the initial stages of acquisition, much less provide them with either encouragement or discouragement about their use of words. Bloom (2000), furthermore, describes a study on children who were unable to speak, so could clearly not receive feedback on their speech, and yet do still develop language normally.

Even when negative evidence is shown to appear, therefore, it is clear that it does not appear very frequently, nor is it culturally universal, contrasting markedly with the learning of words itself, which occurs remarkably quickly and universally, even under the most restricted and deprived of circumstances. Because negative evidence from an external party, such as a parent or teacher, is not able to explain the paradox of word learning, researchers have explored the existence of other constraints within the learners themselves, and it is to these that we turn in the following sections.

3.3.2 Whole-Object Bias

Macnamara (1972) argues that children naturally represent their environment in terms of the objects within it. When learning words, they automatically assume that the new word refers to the whole object, rather than particular parts or properties of the object. For instance, Macnamara (1982), in common with many researchers into language acquisition, describes the development of his child's vocabulary. His son was taught many of the objects involved in the washing and grooming process, such as *soap, toothbrush* and *toothpaste*. The child then associated the word *shave*, which he had not been explicitly taught, to Macnamara's razor. It seems plausible that the child had decided that the unfamiliar word must refer to the only other salient *object* in the event. It is interesting to note further that, in contrast to the very rapid learning of other words, it was very difficult for the child to overcome the association *shave* = RAZOR which he had made[2]. Macnamara hypothesises that this is because changing *shave* to its adult meaning involves thinking not in terms of objects, but actions. The *whole-object bias* leads the child both to create the initial association, and also to resist rejecting it, only doing so very reluctantly.

Macnamara rightly points out that this bias is not just present in children, but is also still present in adults. Firstly, it is particularly apparent when adults teach the names of things to children through ostensive definition. This kind of teaching through pointing at an object and naming it will only work if the teacher can correctly predict the interpretation that the child will give to a new word. In chapter 6, I will show how communication success is considerably improved if the speaker chooses words which are likely to be understood by his interlocutor; the speaker must put himself in the hearer's shoes, and take into consideration the interpretation which a hearer would give to the word. We must be aware, however, that putting yourself in someone else's shoes and deciding what they would think is a sensible strategy only if you are using the same kind of cognitive biases and processes as the other person.

Secondly, Macnamara (1982) gives the example of (adult) learners of foreign languages, who, while beginning to learn the foreign language, use the whole-object bias in order to learn the names of objects as a first step. The language learning process is grounded (Harnad, 1990) on the objects in the world, to which the learner's first words are attached. This particular parallel between first language acquisition and second language acquisition is perhaps not too surprising, and this is, of course, the same scenario as Quine (1960) presented with his imaginary anthropologist hearing *"gavagai"* while the native speaker pointed to a rabbit. Macnamara would contend that the whole-object bias is the very reason why we assume that *"gavagai"* must mean RABBIT.

It has been claimed (Markman, 1989) that the whole-object bias is specifically tailored to word learning, but this is controversial, and Bloom (2001) points out that it is additionally found in a number of non-linguistic domains, such as counting, tracking, categorisation, addition and subtraction. There are further difficulties with the question of what counts as an object, particularly in tricky areas such as meronymy (which deals with part-whole relationships) and temporary attachment. For instance, does the windscreen of a car or

[2]It is interesting that the word which Macnamara's son associated with RAZOR was *shave*, rather than *shaving*. Indeed, in the appendix, Macnamara (1982) records his son as using the word *shaving* to refer to an action a month later than he made the *shave* = RAZOR association. This could point to the child making use not only of the whole-object bias, but also the general linguistic context, plausibly having already discovered that words ending in *-ing* normally refer to actions rather than objects. Bloom (2000) provides a detailed discussion of this phenomenon and other ways in which children appear to use syntactic content to guide them to the meaning of unfamiliar terms.

the tail of a cat count as an individual object, or as a part of another object? In the case of a jockey riding a horse, is the whole rider-horse entity one object or not? Intuitively, we know that all these *could* count as objects in certain circumstances, but equally that, generally, we would not think of them in those terms. But on what basis do we make these decisions? This notion has recently been explored by Spelke (1994), cited in Bloom (2000), who has developed a set of principles which, she argues, we use to decide on whether something is an object. Objects must, according to Spelke, be *cohesive*, *solid* and *continuous*. Spelke further argues that these principles are likely to be innate, and contrast with other principles related to objecthood, such as the fact that unsupported objects fall, which need to be learnt. Bloom (2000) takes this analysis further, making an important distinction between these three of Spelke's object properties, in that solidity and continuity describe the expected behaviour of objects, and as such fit into a prototype based definition of OBJECTHOOD. The principle of *cohesion*, on the other hand, is a more central, classically necessary feature of the definition; things which are not cohesive are not objects[3].

3.3.3 Shape Bias

Useful as the whole-object bias is in explaining how children might bootstrap their language acquisition, it is not a sufficient explanatory tool for the larger problem, and so many researchers have demonstrated additional restrictions, in order to account for more complex facets of word learning. An important discovery in this vein was that children are more likely to categorise new objects in terms of their shape, rather than other perceptual features. Landau et al. (1988) performed an experiment in which they presented children with an unfamiliar object which they explicitly name: "This is a dax". The experimenters then go through a number of test objects, asking the child with respect to each: "Is this a dax?".

In general, children used the new name with objects which were the same shape as the original object, but did not pay attention to size and texture. Perhaps surprisingly, even in cases where the test object was 100 times bigger than the original object, or made of very different substances, the children still chose to name objects on the basis of shape. Armed with these results, Landau et al. propose that children have an innate *shape bias*, or a preference to categorise in terms of shape.

[3]Strictly speaking, it is also possible to override the principle of cohesion in deciding on objecthood; although a child would surely parse a bikini as two objects, the vagaries of fashion determine that it is a single object to adult speakers, despite its being made up of two unconnected parts.

L. Smith (2001) rightly points out that a shape bias is a useful general attentional bias, which secondarily promotes the learning of common nouns, as they typically refer to objects of similar shape. Although the shape bias is clearly not the same as the whole-object bias, they are nevertheless related; organising things in terms of their shape necessitates being able to sort them into objects in the first place, and as shape is such a salient feature of object membership, attending to shape implies attending to objects.

Interestingly, as well as asking them to confirm whether a new object was named with the same term as the target object, Landau et al. also asked children to find objects which were *like* the target object, rather than named using the same term, and found that in this case, children did *not* appear to base their decisions on shape alone, but rather overall similarity was based on the aggregation of a number of perceptual features. This leads them to conclude that the shape bias is used specifically in the domain of word learning.

On the other hand, if there is such a bias, then it is not merely a straightforward 'shape bias' as was originally suggested. Instead, it has been shown that children focus on different properties, which depend both on the linguistic context and the specific properties of the object itself. For example, L. Smith (2001) also reports further studies in which children pay special attention to the texture of the object in addition to its shape, but only if the object appears to have eyes. Soja, Carey, and Spelke (1991) varied the rigidity of objects in their experiments with children's word categorisations. When the named object was rigid, then the word was generalised to things of the same shape, as we have already seen. But when the named object was not rigid, but instead made of something like foam, then the children generalised the word to objects of the same *material*. So if we assume that the shape bias does exist, how does the child know when to apply it, and when to apply what we might facetiously call the non-rigid material bias?

We might argue that the properties of having eyes and of rigidity can be seen as important and plausible in an evolutionary context, in that they point to informative biological distinctions. But L. Smith (2001) also gives details of her replication and extension of the eyed-object study, where objects were additionally shown with trainers. Clearly, an evolved mechanism for naming generalisation which is particularly sensitive to whether a thing wears trainers is totally implausible. Nevertheless, the results showed that children *did* attend to texture when naming objects, but only if the object was portrayed as wearing trainers. It appears that not only are children very good at making generalisations based on objects' properties, but they also seem to learn which of these properties are useful to attend to. Domain-specific learning biases, like the shape bias, might well be used by children in word learning, but it seems possible that the biases themselves may actually be shaped by general development processes, rather than being innately specified.

3.3.4 Taxonomic Bias

Taxonomic organisation is the grouping together of objects of the same type, and is often contrasted to *thematic* organisation, or the grouping of things on the basis of the relations between them. For instance, objects with similar properties (e.g. feathered things which fly) would be grouped together taxonomically, while groupings which are spatially or causally related, like book/table (related by ON) or ball/window (related by BREAK), are organised thematically. Thematic relations are clearly very important in making sense of the world around us, and yet adult categories are not based on thematic relationships (Markman, 1989). The fact that deer leave footprints on the ground allows hunters to track them, but although the deer and the footprints are clearly related[4], they would never be classified as the same type of thing.

Researchers have investigated the variable use of thematic and taxonomic schemes of categorisation in childhood by asking children to sort objects into groups. Broadly speaking, young children (up to 6 years or so) use thematic relations, while older children use taxonomic categories. Different interpretations of these results have caused much controversy, with many researchers (e.g. Vygotsky (1934/1986)) claiming they prove that young children lack the ability to categorise taxonomically, while others would prefer to regard young children as having a preference for, or a heightened interested in, using thematic relationships for this kind of task. Markman (1989) shows that the consequences of children being *unable* to categorise taxonomically would be dramatic and bizarre, as they might have, for instance, ball and window as part of the same category. Markman also claims that an extreme form of thematic categorisation would mean that a category such as ANIMAL would encompass not only dogs, cats, cows, horses and so on, but also things related to them, like a dog lead, cat food, grass and a bridle. But this doesn't happen: children do not find it hard to distinguish cows from grass, and would not entertain the thought that a bridle can run around.

In fact, Markman and Hutchinson (1984) have shown experimentally that children use taxonomic and thematic methods of categorisation differently, depending on whether they are learning words. For instance, when there are no words involved, children will group a car and a car tyre together thematically. When the car is called "dax", however, and the children must try to find another thing named "dax", then the children are much more likely to find the taxonomically related bicycle. Markman and Hutchinson explain these results by proposing a special constraint on word learning. This constraint leads

[4]Indeed, Peirce (1897/1955) would call this kind of linkage *indexical*, one step removed from a basic *iconic* mapping, where the signal resembles the signified, but not arbitrary enough to be a true *symbolic* mapping.

children to suppress their normal, thematic way of interacting with the world, in favour of a taxonomic point of view, if they believe they are learning words. Despite the fact that children like organising things on a thematic basis, Markman and Hutchinson suggest that they have implicit hypotheses about *language*, and in particular word learning, which differ from the way they like to structure objects in their environment.

The *taxonomic bias*, therefore, would appear to be specific to the domain of language. It also works particularly in tandem with the *whole-object bias* (q.v.), so that children interpret unfamiliar words as referring both to whole objects, and also to objects of the same kind. These biases appear to have different timescales, as the whole-object bias is overridden more readily than the taxonomic bias; although children do prefer to allocate unfamiliar words to whole objects, they do eventually learn words for the objects' properties. In contrast, the taxonomic bias appears very strong even into adulthood, so that adult categories too are organised taxonomically and almost never thematically.

3.3.5 Mutual Exclusivity Bias

As we have seen, although the *whole-object bias* and *taxonomic bias* are useful explanatory tools, they are not the whole story. Although it is important for the child as an inductive learner to be able to reduce the number of possible hypotheses, and so begin to learn the meaning of words, they must also be able, in the end, to learn things which violate these biases, which refer, for example, to parts of things and non-shape properties of things. Markman (1989) puts forward a further principle, the assumption of *mutual exclusivity*, in order to overcome the limitations of these biases.

This principle is very straightforward, and states that the extensions of categories are distinct sets which do not overlap. Crucially, however, Markman assumes that mutual exclusivity applies particularly at the privileged *basic* level of categorisation, so that an object cannot be both a dog and a cat. Referring back to figure 2.1, the basic level category would be *cat*, rather than its superordinate *mammal* or its hyponyms *tabby* and *siamese*. Rosch and Mervis (1975) have demonstrated the importance of basic level terms, which are used most frequently, and which are most inclusive. Bloom (2000) states that, at the basic level of categorisation, people judge objects to be more similar, and interact with them in similar ways. Murphy and Lassaline (1997) suggest that the basic level is a compromise between a very specific meaning which can uniquely identify and object, and a very general meaning which is relatively uninformative, as its denotation is too large. It is important to note, however, that the basic level categories like *cat* and *apple* are at the *centre* of a hierarchical structure of meaning, not at the bottom of a chain. Most importantly of all, from the view of word learning, basic level terms are simply those which are used most often by adults, and are among the first words which are learnt by children.

Indeed, it is fair to say that Markman assumes that children only have the basic level of categorisation, initially at least, as it is clear that the mutual exclusivity assumption must be broken in order for any kind of hierarchical semantic structure, such as that described in chapter 2, to emerge at all. Markman does recognise this problem with a strong version of the mutual exclusivity bias, and so she modifies her claim such that children assume that terms are mutually exclusive, until they are presented with overwhelming evidence to the contrary. Assuming we can run with a somewhat vague notion of 'overwhelming evidence', it is at this point that mutual exclusivity is violated, and the child creates the appropriate category.

Markman also discusses a hierarchy of cognitive biases which might apply to the child which is learning unfamiliar words. Although she assumes that the *whole-object* and *taxonomic* biases have a higher priority than *mutual exclusivity*, if an unfamiliar word is applied to an object which *already* has a word attached to it, then mutual exclusivity steps in to avoid the wasteful accumulation of many synonyms to refer to the same thing. The child assumes that each meaning has only one label, and so finds some other salient property to which the term is applied. Mutual exclusivity therefore continues to restrict the number of possible meanings in a very important way, by ruling out those for which the child already has a word. As the child's lexicon develops and the number of words acquired increases, new words have to find new meanings, a process which leads to innovation and the *creation* of novel meaning.

Regier (1995) uses the principle of mutual exclusivity with limited success in a connectionist simulation of the acquisition of spatial terms. Although mutual exclusivity can, to a certain extent, obviate the problem of no negative feedback, by assuming that every positive instance carries with it a set of implicit negative instances for all other meanings, it is only of limited value, as it also necessarily produces false implicit negative evidence. In the basic task of learning names for objects, we have seen that Markman's mutual exclusivity cannot account for the learning of taxonomies or meaning hierarchies. A neat solution to this problem is not so easy to find in Regier's spatial model either. An object can be both *above* and *outside* a landmark, yet every occurrence of *above* will be taken as implicit negative evidence for *outside* and vice versa. In effect, Regier's learning model based on mutual exclusivity succeeds only when the spatial terms actually *are* mutually exclusive (like *inside* and *outside*); in all other cases, the strict implementation of implicit negative evidence means that terms cannot be learnt satisfactorily.

3.3.6 Principle of Contrast

A similar proposal to Markman's mutual exclusivity assumption is the *principle of contrast*, proposed by Clark (1987). According to Clark, any difference in form marks a difference in meaning, and moreover this is crucial in enabling children to make reasoned choices as they try to learn the meaning of words. She makes a number of predictions about language acquisition which follow from the notion of contrast, as detailed below (Clark, 1987, p.10):

1. Children assume words contrast in meaning.
2. Children give priority to known words.
3. Children assign novel words that they hear to gaps in their lexicon, and, to fill such gaps, they coin new words themselves.

Evidence that these predictions do in fact occur comes from a variety of areas. Children are prone to over-extend the meaning of words when they are first learnt, so *dog* initially applies not only to dogs but also to cats and other four-legged animals. But when a child who has over-extended *dog* acquires a new word *cat*, the meaning of which was originally covered by *dog*, the child then carves out a new meaning for the new word, narrowing down the meaning of *dog* so that it does not conflict with the new meaning. This narrowing of meaning also results in the creation of *lexical fields*, or terms for particular semantic areas.

Clark claims that different forms are always allocated a different meaning of some sort by children, so that the one-to-one mapping between form and meaning is maintained. We have already seen how the number of possible meanings to be induced is infinite, so it may not be surprising that the contrasts made by children are not always the same as those in adult language. Clark (1993) gives examples of children establishing a contrast between *duck, bird, chicken*, apparently on the grounds that they swim, fly, and walk. Similar examples are found cross-linguistically: MacWhinney (1985) demonstrates examples of children learning Hungarian, who contrast the nominative and accusative forms of common nouns, but interestingly not in terms of their thematic roles, as in the adult language. Instead, the children have picked up on the fact that adults often use the accusative forms (ending in -t) in questions such as *Kérsz teát?* ('Do you want tea?'), and have generalised this distinction, so that they come to use nominative forms to name things, and accusatives to ask for things.

Children also do indeed give priority to known words, as can be seen when, in the initial stages of acquisition, they reject multiple labels for an object. They do not have an organisation of meanings into hyponyms and superordinates, which allows different levels of labelling, and so cannot accept both *cat* and *animal* as labels for the same object (Clark, 1987). Interestingly, the converse phenomenon occurs in adult learners, who do allow multiple labels, but because of this do not allow exclusive disjunctions of meanings related by hyponymy. For instance, Hurford (1974) marks "John is an American or a Californian" as badly-formed, because *John is a Californian* entails *John is an American*, or *Californian* is a hyponym of *American*. On the other hand, we could hypothesise that young children who reject *animal* as a label for a cat would likely *accept* the similar sentence "Daisy is a cat or an animal", were they to understand logical disjunction, because in their conceptual structure, which rules hyponymy out, *Daisy is a cat* does not entail *Daisy is an animal*.

Furthermore, children have great problems in working out the meaning of nonsense words which are synonymous with words they already know, despite the fact that they acquire new words voraciously (Clark, 1993). In contrast, if the nonsense words are replaced by gaps, then the children can easily find the words which should be inserted. Finding existing words, of course, does not breach the principle of contrast, but, on the other hand, synonyms are explicitly ruled out by it.

This multi-labelling problem also occurs, of course, in multilingual children, who initially reject equivalent terms from a second language when one already exists in another language. The children create a single lexicon, and words in this lexicon, which would be considered translations of each other (e.g. English *no* and Estonian *ei*) in adult language, have different referents in the child's language. Vihman (1996) describes how her son used *no* in many contexts, but *ei* only for self-prohibition. Clark (1987) hypothesises that the point at which synonym 'doublets' are allowed in the bilingual child's lexicon may coincide with the time at which it can distinguish the two languages phonologically; as they realise that they are dealing with two systems, they can then naturally accept words in both systems. Within each system, the Principle of Contrast still applies, but the children can acquire a high number of doublets (Clark, 1993).

On the other hand, Deuchar and Quay (2000) report the results of a comprehensive study of a Spanish/English bilingual child, which appear to clearly contradict Clark's claims for how the principle of contrast interacts with the acqusition of two languages at the same time. Deuchar and Quay show both that the child in their study had equivalent terms, for example *bye/* **tatai, más**/*more* and **zapato**/*shoe*[5], and also that at no stage does she appear to reject equivalent doublets. More damagingly for Clark, these doublets occur from an early age, even under eighteen months, and certainly well before the age of two years, which has been suggested as the time at which children appear to be aware of phonological differences (Vihman, 1996).

Deuchar and Quay (2000)'s study of bilingual acquisition does indeed pose problems for the principle of contrast, which Clark (1993) has already anticipated when she appeals to an analogy with linguistic *registers* or a situation of *diglossia* (Fasold, 1984). Diglossia occurs when people use two distinct codes or registers of language in two completely different situations. Prestigious, or high registers are characteristically used in formal, religious, and legal contexts, while low registers are used in informal, casual contexts. Diglossia is a very widespread phenomenon. There are, for instance, a large number of low varieties of Arabic across North Africa from Morocco to Egypt, through the Middle East and the Gulf States, which are used in most informal, everyday contexts and are frequently mutually unintelligible. In order to be understood across the Arab world, children are taught to read and write the higher variety of (Modern) Standard Arabic. Still more prestigious, and regarded even as sacred, is the highest form, Classical Arabic, found in the Qur'an[6] and in political communications. Diglossia also occurs historically, and with varieties that are clearly different languages, such as in England in the three centuries following the Norman Conquest, when the high register Norman French and the low register Middle English existed together.

Clark (1987) claims that words from different registers or varieties are not true synonyms, because of the different functional situations in which they are used. Analogously, the bilingual child's synonyms are, according to Clark, not true synonyms, but merely translation equivalents like *horse* and the Czech *kůn*, but Deuchar and Quay point out that this assumption holds only if we additionally agree that the bilingual children know that they are learning two languages. Because of the nature of the developmental process, it is actually very difficult to draw any conclusions at all about how many linguistic systems are concurrent. Phonological and syntactic evidence for a distinction between Spanish and English, such as the distinct use of particular language-specific phonemes or of grammatical agreement categories, arrives much later than the occurrence of the doublets in the child's lexicon, casting doubt on Clark's assertion that synonyms are only possible after the original language system has been divided into two separate systems.

[5]The doublets are given in order of acquisition by the child, with the Spanish terms in **bold**.

[6]Because of the religious significance of Classical Arabic, many speakers of Arabic regard Classical and Standard Arabic as one and the same, and the local varieties as impure, corrupted languages.

The problems of pinning down what is actually happening in bilingual acquisition notwithstanding, the principle of contrast also predicts that children will assume that an unfamiliar word will refer to a gap in their lexicon, a meaning without a word, and that, in production, they will construct words to fill these gaps. The original study which looked at the designation of unfamiliar words was by Carey and Bartlett (1978), who introduced children to a new word by asking them to find a *chromium* cup or tray. The new word was contrasted with a familiar colour term (either red or blue, depending on the object), and most of the children decided, in accordance with the principle of contrast, that *chromium* must mean the unfamiliar colour. Interestingly, most of them asked for some confirmation about which object the experimenters wanted, having clearly marked *chromium* as an unfamiliar word.

When children want to express meanings for which they have no word, some process of innovation occurs. Clark (1993) shows that children's preferred forms in word coinage are those options which are productive within their speech community, so they learn how to make words as part of the language acquisition process. Children can find smaller units inside larger words, and after exposure to several instances of these smaller units, can map some meaning to them. The small form-meaning unit can then be used productively by the child in the creation of innovative forms.

This kind of analysis and decomposition of signals into smaller units of meaning is used by Kirby (2002) in a computer simulation of the emergence of compositional syntax. Agents in the simulation, whose role is analogous to children learning a language, take advantage of coincidental matches between parts of utterances and parts of meanings to create general rules. When the agents come to attempt to produce utterances for meanings, these general rules can necessarily produce more utterances than idiosyncratic, holistic rules, and so the utterances created by general rules are more likely to be produced, and maintained in the population.

Some kinds of meaning-form pairs are easier for a child to induce than others[7], and indeed there are large differences between languages in the degree of transparency in areas such as nominal and verbal inflectional morphology. Hungarian, for instance, has a very rich set of locative suffixes which are added to all noun phrases, some of which are shown in table 3.1 attached to the root *ház* (house). We can divide Hungarian locatives easily into three main sub-groups, representing the broad spatial relationships denoted by IN, ON

[7]In simulations such as those by Batali (2002), Brighton (2002), Kirby (2002), the meanings are not actually induced at all, but are instead given to the learning agent. We will investigate an alternative to this approach in the communication system described in chapter 6.

	IN	ON	AT
AT	*a házban* in the house	*a házon* on the house	*a háznál* at the house
TO	*a házba* into the house	*a házra* onto the house	*a házhoz* to the house
FROM	*a házból* from in the house	*a házról* from on the house	*a háztól* from the house

Table 3.1: Some Locative Expressions in Hungarian

and AT, each shown in a column of the table. Each column is further divided into three rows, representing location at, motion towards, and motion from.

In Slavic languages such as Serbo-Croat[8], on the other hand, these and similar locative meanings are marked by both prepositions and endings on the noun. Many of the prepositions take more than one case, depending on the meaning, and the same case can also occur with different prepositions, again dependent on the meaning; some examples of these are given in table 3.2, using the equivalent noun *kuća*, which also means house (Schmaus, 1961).

It is clear that there is not such a straightforward mapping between form and meaning in this area of Serbo-Croat as there is in Hungarian. Does this difference in the form-meaning mapping make any difference to the children learning the languages? Slobin (1973) shows that it does: Hungarian/Serbo-Croat bilingual children produce the Hungarian locatives correctly at the age of two years, while the same children leave out prepositions and use cases inconsistently when using Serbo-Croat, not mastering the less straightforward system until around the age of five.

The transparency of the Hungarian system, with roots always being followed by the locative markers, and with small parts of words having distinct and consistent meanings (in table 3.1, for example, all the endings in the IN column begin with *'-b-'*, and all the endings in the FROM ROW end in *'-ól'*), means that children can generalise successfully more easily, using the principle of contrast, and innovate to fill gaps in their lexicon. The principle of contrast is less immediately useful in learning Serbo-Croat locatives, because

[8]Since the break-up of Yugoslavia and the concomitant wars during the 1990s, it has been politically expedient to regard Serbian and Croatian as separate languages, but linguistically the differences between them are minor, being restricted in the main to individual vocabulary items and (most notably) the script in which they are written.

Same preposition, different cases:

u kuću	Accusative	into the house
u kući	Locative	in the house

Same case, different prepositions:

kod kuće	Genitive	at home
blizu kuće	Genitive	near the house

Table 3.2: Some Locative Expressions in Serbo-Croat

there is no straightforward mapping between the relevant morphemes, and so no easy way to make contrasts between expressions; children consequently make more mistakes when producing innovative sentences in Serbo-Croat.

The principle of contrast, therefore, and more specifically the assumption of mutual exclusivity, appears to be a very important process in the acquisition and maintenance of language. Its essence, that every change in form is reflected by a change in meaning, is simple and powerful in helping learners escape from the Quinean paradox of meaning induction, as I show experimentally in chapter 9.

3.3.7 Word Learning without Specific Constraints

In the previous sections, we have investigated many specific constraints on word learning, which have been proposed to account for experimental evidence regarding children's achievements in language learning. Bloom (2000, 2002) and Tomasello (1999, 2001b) separately propose alternatives, in which children do not have to be endowed with constraints which are specific to word learning. Bloom (2000) makes it clear that this does not return us to the Quinean paradox of an infinite set of possible meanings, because children clearly *are* constrained somehow in the meanings which they will consider. The main problem, according to Bloom, is that these constraints do not need to be specific to the domain of language learning, and are more profitably and parsimoniously explained in terms of general ideas about how children think and learn. Tomasello (2001b) suggests also that children's learning of words does not occur via specific hypothesis testing, but instead is part of a general, social-pragmatic development of cultural skills and conventions:

> "Children learn words as an integral part of their social interactions with other persons, an important part of which are their attempts to understand what adults are trying to get them to do and their attempts to get adults to do things." (Tomasello, 2001b, p.136)

Tomasello and his collaborators have, over the last couple of decades, published details of many experiments which attempt to show that the word-learning constraints we have been looking at in this chapter are not necessarily the best explanation of how words are learnt. In all cases, the experiments are set up so that children and adults are interacting, and unfamiliar words are dropped into the conversation naturally. Pragmatic cues are provided to the children to see whether they are sensitive to them, and attempts are made to neutralise the possible confounding impact of the word learning constraints themselves.

For example, Tomasello and Barton (1994) set up a study in which the experimenter would say to the children: "let's go find the toma". Both then approached five buckets, all filled with different, unfamiliar objects. On some occasions, the adult went straight to the appointed bucket, took the target object, and gave it to the child; on others, she searched through the buckets, extracting objects, scowling at them and replacing them, until she found the target object, again giving it to the child thereafter. Later, the child was shown the object, and asked its name; under both circumstances, the children learnt the new word equally well. Tomasello and Barton suggest that the children could use neither cues such as "the object the adult is looking at while saying the word", but must instead have understood the adult's intention to find a particular object, and have been able to evaluate the adult's fulfilment of *their* goal. A similar study by Akhtar and Tomasello (1996) used variations on the same experimental theme, except that one of the buckets was very distinctively different from the others, and the children were primed with non-linguistic games so that they would discover which object was in the distinctive container. The experimenter would then try, and fail, to get into this container, in which the child knew the mystery "toma" was located. The experimenter could therefore never actually find the object, and only ever showed disappointment at not doing so, and yet, even despite the fact that the goal of finding the object was never fulfilled, the children still successfully learnt the name of the hidden object, showing that children use a flexible and diverse set of strategies to work out the communicative intentions of their interlocutors.

Carey (1978) proposes the notion of *fast mapping*, whereby learners accurately learn the meaning of a word based on hearing them as little as a single time. We have already discussed her experiment with children learning colour terms with respect to the principle of contrast (Carey & Bartlett, 1978), but the children's learning that *chromium* means olive-green is also evidence in favour of fast mapping. After six weeks in which the word was never again used, the children were shown an olive-green object and asked to describe it. Most of them remembered the word they had heard just once, and the others all used a different colour name (e.g. grey, brown), which they had not yet stabilised in terms of reference.

Bloom (2001) describes an extension of Carey and Bartlett's experiment, in which both children's and adults' capacity for fast mapping was tested (Markson & Bloom, 1997). The subjects were exposed to an unfamiliar word *koba*, which referred to one or more unfamiliar objects. As a separate experiment, they were also exposed to another object, which was referred thematically, either by the use of a phrase which referred to another linguistic entity ("my uncle gave these to me") or by explicit visual demonstration ("this goes here"). Across all age groups, the subjects remembered which object was referred to by *koba* over half the time. Interestingly, they also did so for the linguistically presented facts, but did *not* do so for the visually presented ones. Bloom (2000, 2001) uses these findings to argue that fast mapping is a general purpose mechanism, not used just for word learning; not only does the process occur in other contexts than word learning, but adults have the ability as well as children, so it is not part of any specialist language learning apparatus which disappears after the *critical period* (Hurford, 1991) for language acquisition.

Instead, Bloom (2000) argues for a combination of the different, general, cognitive adaptations we have discussed: an ability to see the world in terms of *objects, events, relations, kinds* and *individuals*; the ability to generalise (and, crucially, to make the right generalisations); an insight into the intentions of others (also known as a theory of mind), and an understanding of what they are referring to; an understanding that some categories are *not* reducible to their observable features (see the discussion on naïve essentialism and the theory theory in section 2.3.3); and the ability to count. Interestingly, given his co-authorship of the seminal article which argued the case for the natural selection of the Chomskyan Language Acquisition Device (Pinker & Bloom, 1990), Bloom now argues that there is *no* need to posit the existence of any specific 'language-learning module', at least in terms of learning words. Related to the use of general cognitive faculties is the use of heuristics in order to get round the problem of an infinite search space. Gigerenzer and Todd (1999), for example, show how the use of simple heuristics can often provide simple and elegant answers to such problems, without the need for costly and specialised cognitive architecture.

Bloom (2001)'s alternative proposal twists the cognitive biases we have previously discussed on their heads, saying that we perceive a whole-object bias for word learning because whole objects are salient, and count nouns refer to kinds of individuals; a shape bias because words refer to kinds of objects, and we often categorise on the basis of shape. These are very general cognitive heuristics, which have nothing explicitly to do with word learning. In this respect, Bloom's argument is not so very far from L. Smith (1999)'s position that although there are specialised learning mechanisms like the shape bias, these are constructed from general learning processes, in an analogous fashion to the specialisation in cell development which leads to the differentiation between liver cells and brain cells, for instance. Both argue that the source of the biases is domain-general rather than domain-specific, but Bloom would go further, and deny that there is any benefit in saying there is a bias at all.

Finally, support for the fast mapping theory of word learning may be found in a recent neurological study by van Turennout, Ellmore, and Martin (2000), in which they provide evidence of long-lasting plasticity in different parts of the brain, notably the occipitotemporal cortex, left inferior frontal and left insular cortex. Interestingly, the left inferior frontal is considered part of Broca's area, a part of the brain well-attested in syntactic and phonological processing (Deacon, 1997)[9]. In their study, van Turennout et al. (2000) suggest that initial naming of an unfamiliar object, as in fast mapping, is dependent on Broca's area, but that during the repeated retrieval of an object's name, when the process becomes more automatic, Broca's area is used increasingly less, with a corresponding increase in activity in the left anterior insula.

3.4 The Nature of the Learning Task

We have discussed in some detail the problem of meaning induction demonstrated by Quine (1960), and many of the proposals which seek to explain how children overcome the problem with such ease. Related to this problem is the more complicated issue of what learners actually do learn when they learn the meaning of a word. In other words, what kind of mapping is the learning task they perform?

Although all languages map words onto meanings, and although we must assume that the set of possible meanings available to humans is universal and potentially accessible to speakers of all languages, the way in which languages divide up this semantic space is

[9]Although Broca's and Wernicke's areas are the most commonly mentioned areas of the brain which have a 'specialisation for language', there is in fact a vast literature claiming more and more different areas with specialist language functions. Beaken (1996), for instance, lists almost two dozen different areas which have been proposed. Despite all this work on linguistic neurology, we cannot reliably identify the area of the brain containing either the Language Acquisition Device or the 'language controller'; in neither case can we pinpoint any small area of the brain, which, if destroyed, would inhibit language competence.

very different. The following brief examples show how even in some of the areas which humankind has most in common, remarkable differences can be seen in the semantic organisation of different languages. For example:

- We have already seen in sections 2.3.3 and 2.5.1 how Kalam-speakers classify the wildlife in their surroundings, and how it is impossible for an English-speaker to translate a word like *kmn*, without resorting to an enormous explanatory paraphrase, or, worse, a list of all the animals which are covered by the word and another list of those which are excluded;
- Wierzbicka (1997) explores in great detail how even apparently basic cultural categories such as the notion of 'friendship' differ dramatically in three related European languages: English, Polish and Russian.

But differences in classification systems which deal with the animals found around where you live, and with cultural concepts, however universal they might seem, are not necessarily so strange. More surprising is that differences in the division of semantic space are found even in areas which are unquestionably universal:

- A simple concept such as *my brother* can be translated straightforwardly into Hungarian in many different ways: *öcsem*, my younger brother; *bátyám*, my elder brother; *fívérem*, literally a 'son of my blood'; and most generally *testvérem*, my sibling, the latter being used much more commonly than its English equivalent might suggest.
- It is hard to think of a more common human experience than the human body itself, and if meaning is based on experience, we might imagine that classification and categorisation of body parts should be universal. Although I know of no language which does not have distinct words for body parts such as *head*, most Slavic languages such as Czech use a single word (*ruka*) for the whole of the arm, including the hand, and similarly a single word (*noha*) for the whole of the leg, including the foot; it is interesting and instructive to read in a English dictionary for Czech speakers that 'arm' is glossed as '*ruka* above the wrist', and 'hand' as '*ruka* below the wrist'.

  In other languages, such as Hungarian and Albanian, a distinction is made between *arm* and *hand*, but not between *leg* and *foot*, where one word is used for both

Table 3.3: Language-specific differences in the categorisation of body parts

<table>
<tr><th>Family</th><th>Language</th><th>'Arm'</th><th>'Hand'</th><th>'Leg'</th><th>'Foot'</th></tr>
<tr><td rowspan="3">Indo-European</td><td>Albanian</td><td>krah</td><td>dorë</td><td colspan="2">kâmbë</td></tr>
<tr><td>Czech</td><td colspan="2">ruka</td><td colspan="2">noha</td></tr>
<tr><td>Dutch</td><td>arm</td><td>hand</td><td>been</td><td>voet</td></tr>
<tr><td rowspan="2">Finno-Ugric</td><td>Finnish</td><td colspan="2">käsi</td><td colspan="2">jalka</td></tr>
<tr><td>Hungarian</td><td>kar</td><td>kéz</td><td colspan="2">láb</td></tr>
<tr><td rowspan="2">Bantu</td><td>Swahili</td><td colspan="2">mkono</td><td colspan="2">mguu</td></tr>
<tr><td>Xhosa</td><td>ingalo</td><td>isandla</td><td>umlenze</td><td>unyawo</td></tr>
</table>

concepts. Differences in the classification of limbs are actually reasonably well-attested, and as we can see from table 3.3, occur both across and within language families.

The task facing the human language learner, therefore, is not as straightforward as simply mapping between words and meanings. As we saw in chapter 2, we must also explain where the concepts come from, as well as how they are linked to words. The massive differences in meanings across languages, which we have only touched upon in this section, seem to imply that at the very least, there must also be some kind of mapping between word meanings and some units of conceptual or semantic space.

3.4.1 Language-specific Categorisation

And yet we find that the semantic units from which this conceptual space must be built are themselves apparently not made up of any easily accessible cognitive primes. Brown (2001) and de León (2001) give very interesting accounts language-specific semantics in the spatial terms of the related Mayan languages Tzeltal and Tzotzil, which are spoken in the Chiapas highlands of Mexico. These languages use a system of describing location which is completely foreign to an English speaker. Speakers of Mayan languages appear to regard the whole world as if it were tilted down northwards, so they speak of the 'uphill' end of a table. Levinson (2001) gives the following example from Tzeltal:

(3.1) pachana bojch ta y-anil te kartone
bowl.put.CAUSE.IMP gourd.bowl at its.downhill the cardboard.DEIC
'put the bowl behind the box'

In order to translate the English concept *behind*, Tzeltal speakers use the term *y-anil*, which literally means 'its downhill side'. Although the two sentences can be good translations of each other, this hides a great difference in the conceptual systems of the language. Levinson (2001) draws a distinction between the following three strategies for specifying the spatial position of an object:

intrinsic, in which the object is located in a domain specified by the Ground (Talmy, 2000) object.

e.g. *'behind the box'*

relative, in which the object is located relative to the speaker.

e.g. *'to the left of the box'*

absolute, in which the object is located according to a fixed, geographical frame of reference.

e.g. *'to the north of the box'*

As we can see, the English preposition *behind* is, using Levinson's terms, an *intrinsic* direction term, which locates one object in a domain specified by the object which serves as the argument of the preposition. By using *behind* in a translation of *y-anil* as in example 3.1, the speaker means that the box should end up between herself and the bowl. If the positions of speaker and bowl are reversed, the English speaker can still use *behind* to refer to the situation, because their positions relative to the Ground object (the box) are the same. In Tzeltal, on the other hand, *y-anil* is a cardinal or absolute direction term, in which the frame of reference is fixed according to the local landscape. In the second situation, with the positions of speaker and bowl reversed, a Tzeltal speaker could now not use *y-anil*, but would instead be obliged to use *y-ajk'ol*, meaning its uphill side (Brown, 2001).

Clearly, spatial concepts in Tzeltal are very different to those in English, and translation between the two is not straightforward, as much additional contextual information is required for an accurate translation. How, then, do children manage to build such different conceptual systems and induce appropriate meanings for the words they hear? Brown (2001) shows that semantic units which had previously been considered as universal building blocks, such as VERTICAL, do *not* provide Tzeltal children with a set of concepts onto which they can map words as they induce their meanings from context; instead, they appear to develop the concepts themselves through the process of learning words in context.

Choi and Bowerman (1991) provide further evidence of language-specific categorisation in children, in their study of English- and Korean-speaking children categorising spatial events in spontaneous speech. Choi and Bowerman focused on English and Korean because of the notable differences between the languages in categorising actions relating to the positioning of one object relative to another. English is a 'satellite-framed' language[10], in which the path of the action is determined by the satellite, and Korean is a 'verb-framed' language, in which the path is determined by the verb (Talmy, 2000). For instance, Choi and Bowerman show how English makes a fundamental distinction between contact with an external, supporting surface (putting *on*), and putting into a container (putting *in*). This distinction is unknown in Korean, and instead a distinction is drawn between putting two objects into a close-fitting, interlocking relationship (*kkita*), and putting two objects into a loose-fitting relationship (*nehta*). For instance, putting a ring *on* a finger, a top *on* a pen, a cassette *in* its case, and closing a filing cabinet drawer are all covered by *kkita*, while putting an apple *in* a fruit bowl or a quoit *over* a pin are covered by *nehta*.

Choi and Bowerman's most important finding with respect to the way that children learn to categorise spatially was that the children in different language communities categorised spatial events *language-specifically*; they did not use any universal or basic set of semantic concepts, but instead the English-speaking children used the categorisation scheme of adult English, and the Korean children the different system of adult Korean. Bowerman and Choi (2001) suggest persuasively that this language-specific learning has already started by the second half of the second year of a baby's life, and that the children's sensitivity to their language-specific distinctions begins to develop in comprehension before production.

These results, then, pose difficulties for accounts of acquisition which rely on universal spatial primitives or units of conceptual space. An important insight which may shed some light on how to explain this problem is that many apparently diverse linguistic categories can actually form continua. Schlesinger (1979) shows how the comitative case, which expresses TOGETHER WITH and the instrumental case, which expresses BY MEANS OF can be regarded as the two ends of the same conceptual continuum, which just happen to be usually expressed using the same preposition in English, as can be seen in his ordered list of ten simple English sentences, reproduced below:

[10] A *satellite* is a non-nominal complement to a verb root, like the verbal particles *in, out* in 'He went *in/out*' and the German or Hungarian verbal prefixes with similar meanings ('Er ist *hinein/hinaus*gegangen','*Be/Ki*ment').

(3.2)
a. The pantomimist gave a show with the clown.
b. The engineer built the machine with an assistant.
c. The general captured the hill with a squad of paratroopers.
d. The acrobat performed the act with an elephant.
e. The blind man crossed the street with his dog.
f. The officer caught the smuggler with a police dog.
g. The prisoner won the appeal with a highly paid lawyer.
h. The Nobel Prize winner found the solution with a computer.
i. The sportsman hunted deer with a rifle.
j. The hoodlum broke the window with a stone.

Schlesinger presented the set of sentences to speakers of languages which do not use the same form for the comitative and instrumental meanings, and found that although different languages divided this continuum at different points, the *ordering* of the continuum itself was never broken. A language like Swahili, for example, used the preposition *na* in sentences a-f, and the preposition *kwa* for sentences g-j, but he found no language which used the same word for sentences c-e and also h-i, for instance.

Bowerman and Choi (2001) describe a previous study (Bowerman & Pederson, 1992) in which a similar spatial continuum from support (cup *on* a table) to containment (apple *in* a bowl) was demonstrated. Again, languages divided this continuum up in different ways, and again it is interesting to find that they always maintain the integrity of the continuum itself. Dutch, for instance, uses *op* for support and adhesion (a plaster *on* a leg), *aan* for attachment (a picture *on* a wall) and suspension (an apple *on* a twig), and *in* for containment, while Spanish uses *en* for all these relationships. Bowerman and Choi hypothesise from these studies that children might use similarity gradients to guide semantic learning, using these kind of continua to generalise systematically.

These detailed studies show us, therefore, that an accurate model of word learning needs to include not only an associative module which links words and meanings, but also a mechanism for the construction of meaning, created in response to the learner's experiences in its environment, and a way for the learner to work out which of the meanings it *can* create, *should* be used in the language it is learning. It looks uncomfortably likely that we need to expand the learning task yet again, to include not only mappings between words and word meanings, and word meanings and universal concepts, but also another level of mapping between universal concepts and cultural semantic parameters.

3.4.2 Linguistic Relativity

The scenario outlined by Bowerman and Choi (2001), Brown (2001), and de León (2001) can be seen as giving support to the weaker variant of the (in)famous Sapir-Whorf hypothesis, known as *linguistic relativity*, in which language is held to influence thought directly. The stronger version of the hypothesis, or *linguistic determinism*, stating that language constrains the thinking of people has been regularly, and mercilessly, attacked as an object of ridicule, for instance by Pinker (1994), yet despite Whorf (1956)'s own rather mystical observations, the recast, weaker version of the hypothesis appears relatively sensible given all the evidence we have seen about the language-specific differences in semantic structure, and indeed it is often unwittingly supported by many people who claim to disagree with it. The Sapir-Whorf hypothesis remains a contentious topic of debate, yet still far from conclusively proven or discounted. Indeed, Cowan (1997) describes the grammar of *lojban*, a language which was designed specifically to test the hypothesis, by allowing the full expressive power of a natural language but with differences in structure and with a grammar based on the principles of logic. For the moment, however, until a diverse international community of lojban speakers emerges, we must investigate the influence of language on concepts by looking at existing human languages.

Firstly, let us look at the popular and relatively successful movement to rid languages of sexist terminology (Lakoff, 1976; Maggio, 1989) such as the generic use of 'he' and words such as 'chairman', 'policeman' and so on. Even in this realm of promoting linguistic equality, there are intriguingly different strategies for the coinage of replacement words which appear to depend explicitly on the coding of gender within the language. We can consider two main types of gender system, following Corbett (1991), as follows:

semantic gender systems, in which nouns are assigned to a class based on their meaning. For instance, Tamil has a strict semantic system, in which nouns are assigned gender based on the sex of their referents, as does Modern English, where *man* is masculine, and is referred to with the pronoun *he*, while *woman* is feminine, and all inanimate nouns are neuter[11]. Also included here are also predominately semantic systems, such as Dyirbal, in which each gender has a clear semantic basis, but there are numerous exceptions because the bases are not mutually exclusive;

[11] In English, however, there is a degree of confusion about the gender of nouns denoting children and animals, and a possible exception with 'ship', which is often feminine.

although fish are generally in class I as non-human animates, poisonous fish are in class II, perhaps due to association with fire and other harmful substances.

formal gender systems, in which nouns are assigned to a class based on morphological and phonological reasons, which may have no relation to the sex of the noun's referent. In Old English, for example, two of the words for 'woman' were not in the feminine gender: the ancestor of our modern word *wīfmann* was masculine, while another word *wīf* was neuter, as is also reflected in its German cognate *das Weib*.

In languages with a natural gender system, like English, the mechanism for rooting out sexist terminology is to create gender-neutral forms: 'chairperson' replaces 'chairman' and 'police officer' replaces 'policeman'. These replacement forms can be referred to with masculine, feminine, or often generic gender-neutral pronouns such as 'they'. In languages with grammatical gender, on the other hand, gender-neutral forms are often not permitted in the language, and with this avenue excluded, innovators instead resort to coming up with feminised forms of sexist terms, which of course are just as sexually exclusive as the original offending term. In French, for instance, new coinages include *écrivaine* and *auteure* to describe explicitly female writers, because the standard terms (*écrivain, auteur*) are grammatically masculine and cannot be made gender-neutral (Pauwels, 1999). It is hard to see how these two diametrically opposed strategies can be reconciled without acknowledging that the main pressure for choosing one over the other comes from the explicit coding of gender within the language itself, thus providing clear supportive evidence for Sapir-Whorf. Indeed it is reasonable enough to ask what the *purpose* of the 'linguistic equality' movement is, if language does *not* have any influence on the thoughts of its speakers?

One way in which people have sought to investigate linguistic relativity is to choose a small, well-defined domain, and then look at how it is organised in various languages. The most famous of these studies, undertaken by Berlin and Kay (1969), discovered, after investigating speakers of twenty different languages, that basic colour categories were universal, that there was a great deal of agreement on the examples, or focal points referred to by each basic colour term, and moreover that there was a specific hierarchical order to the emergence of these colour terms. Basic colour terms are, in Berlin and Kay's terminology, both general and salient, that is, they apply to diverse classes of objects, and are readily available to most speakers of the language. This has been taken as compelling evidence against linguistic relativity and in favour of a nativist specification of semantic categories, though given the well-understood nature of the human visual system (see Belpaeme (2002) for a detailed discussion), it would be better to think of it as an exploration of the innate visual biases which lead us to divide up the colour space into semantic categories.

Although Berlin and Kay's claims are widely reported, their methodology has been the subject of much criticism, and indeed they are often accused of making the data fit the hierarchical sequence which they proposed. In particular, Sampson (1997) has investigated the details of the data on which their conclusions were drawn, and found a number of problems which cast doubt on their reliability. Firstly, the data was gathered by students on one of their courses, who chose a language, learnt about its colour terms as best they could, and reported their findings as coursework. In a number of cases, the information reported by the students is unsurprisingly wide of the mark. Sampson reports a particularly entertaining example where Berlin and Kay report colour terms in Ancient Greek, but somehow fail to find the very common word *melas*, which means BLACK and is even now used relatively productively in deriving scientific English words from classical roots (e.g. *melanoma*, a cancer consisting mainly of black pigment). Unfortunately, however, BLACK and WHITE are the first two terms on the colour hierarchy, and the authors therefore are in need of a word. They settle on the obscure *glaukos*, which actually had little or no colour reference in Ancient Greek[12], though it denotes a blue-green-grey colour in Modern Greek. Sampson (1997) also shows how Chinese loan colour terms are excluded from the Korean set of colours, which then fits the hypothesis, but they are included in the Vietnamese list, which would not fit without them (and is only a marginal case even with them).

There is also much criticism of an apparent cultural bias towards American English in Berlin and Kay (1969)'s work, and an inherent assumption that the colour categories in American English are at the 'highly-evolved' end of an evolutionary scale. There are in fact more basic colour terms than the eleven they name; Russian, for instance, has at least twelve (Goddard, 1998), and there may even not be any basic colour term present in all languages, nor are the best examples of each category quite as predictable as Berlin and Kay would have us believe (Dedrick, 1998).

[12]Sampson (1997) appears to be exaggerating with his claim that *glaukos* had no colour reference at all, as Liddell and Scott (1980)'s standard Greek-English lexicon shows its primary meaning as 'gleaming, glancing, bright-gleaming', with a secondary meaning of 'pale green, bluish-green, gray' in the restricted field of reference to the colour of olives, willows and vines.

3.5 Summary

In this chapter, we have investigated the problem of how children learn the meanings of unfamiliar words. We have seen that there is a very real problem of meaning indeterminacy if we assume simple inductive learning, because there is always an infinite set of possible meanings, no matter how much evidence is collated.

To try to get round this problem, many innate cognitive biases, which allow the learner to reduce the set to a finite one, have been proposed and explored. In particular, we have seen experimental evidence for the whole-object bias, the shape bias, the taxonomic bias, the mutual exclusivity assumption and the principle of contrast. We have seen further how different languages divide up semantic space in different and incompatible ways, and seen that a complete account of the learning of words must include not only the learning of a mapping between words and their meanings, but also a way to work out what kind of semantic organisational structure is used by the language being learnt. Semantic categories are not shaped directly by conceptual biases, but only in interaction with this semantic organisational structure.

In the next chapters, I will describe my model of experience-based meaning creation and communication, and then go on to investigate how the inclusion of cognitive biases such as those discussed here can affect the conceptual structure of agents and their success in developing a mutual communication system.

CHAPTER 4

The Representation and Creation of Meanings

"It is difficult to design and motivate empirical studies on concept acquisition without first committing oneself to a set of assumptions about what concepts are and how they are represented." (Keil, 1992, p.25)

4.1 Introduction

In chapter 2, we explored the nature of meanings and how concepts can be acquired, then in chapter 3 we investigated the particular problem of how learners can learn the meanings of unfamiliar words. In this chapter, these two strands will be linked to the wider field of evolutionary linguistics as discussed in chapter 1, as I take a look at recent simulations of the evolution of aspects of human language, and in particular at the models of meaning representation and meaning creation which have been put forward in the literature, and the model which I will adopt for the simulations in this thesis.

The linguistic competence of a language user falls naturally into three different, but mutually connected major subsystems: phonology describes the linguistic coding of the signals which are heard and uttered, semantics describes the coding of the meanings which are expressed and understood, and syntax can be regarded as the mapping between phonology and semantics. Although many linguistic theories choose to ignore or gloss over some of these subsystems, it is clear from the last four decades' work in linguistics that a comprehensive theory of language must address all three subsystems, as well as the interactions between the three. Keil's concerns in the epigram at the start of this chapter with respect to empirical studies with children are no less true when designing models for investigations using agents[1]. In order to be implemented on a computer, all three subsystems of language must be represented symbolically, and in different ways, so that they can then be interpreted by researchers as in some way 'phonological', 'semantic' or 'syntactic'. The way in which these systems are implemented is heavily dependent on the theoretical assumptions of the designers of the simulations, and it is to these methods that I now turn my attention.

In this chapter, I explore in detail representations of meaning and mechanisms of meaning creation which have been put forward in evolutionary linguistic simulations, and then, building on the conclusions I draw from this, in chapter 5, I present my own model of semantic representation and meaning creation, which is used in the experiments in subsequent chapters. In more detail, section 4.2 is a discussion of the various semantic representations which have been used in recent simulations of aspects of language evolution in a little detail, discussing in particular how they relate to the famous dichotomy of meaning between *sense* and *reference* (Frege, 1892), and investigating the assumptions which have been made about how meanings are acquired and how they spread through a population of agents. In section 4.3, I then move on to look at the same simulations from the point of view of meaning creation, investigating the mechanisms which have been put forward to adapt internal semantic representations, and will suggest a suitable method for grounded, individually created semantic representations.

4.2 The Representation of Meanings

4.2.1 Predicate Logic

In the models of Kirby (2000, 2002), Hurford (2000) and Batali (2002), meanings are based on variant representations of first-order predicate logic, probably the most widely used knowledge representation language for describing the semantics of both simple propositions and fairly complex facts about the world which are derived from the simpler facts by standard formal rules of inference.

In the earliest of these evolutionary models, Kirby (2000)'s meanings each have three attributes, as shown in 4.1. He glosses them with standard linguistic theory as *agent*, *patient*, and *predicate*, while rightly emphasising that it is important to remember that

[1] In this thesis, I am not using the term *agent* in its usual linguistic sense of the logical subject of a transitive clause (see Song (2001) for an exposition of how the agent role is realised in different languages), but instead in its very common artificial intelligence sense, where it simply means a 'simulated individual'. Under this umbrella term I include all simulated individuals, whether they exist only inside a computer or are physically implemented as robots.

these glosses do not exist in any way inside the simulations; they are simply a mnemonic which helps us to understand the structure of the meanings, which in effect are presented as a straightforward representation of a sentence containing a transitive verb with a subject and object. Kirby's three attributes are further classified, again in accordance with standard practice in this field, so that *agents* and *patients* are classed together as subsets of *objects*, while *predicates* are classed as *actions*. There is, therefore, a very precise typed structure to the meanings in Kirby (2000)'s semantic model; a particular object can appear in either of the two attributes which are available to it, namely as agent or patient[2], but it is impossible, for instance, for an action to occur as either agent or patient, or for an object to occur as the predicate.

$$\text{Meaning} = \; < \overbrace{\text{Agent} \;\;,\;\; \text{Patient}}^{\text{Objects}} \;\;,\;\; \overbrace{\text{Predicate}}^{\text{Actions}} > \tag{4.1}$$

In Kirby (2002)'s extension of this research, which focuses on the emergence of compositionality and recursion, the concepts are similar, as shown in 4.2, although there is a crucial extension. There are now two types of predicates: the first identical to that shown in 4.1; the second is a new type of predicate, which instead of an object as its second argument, takes another meaning representation. There are no further restrictions on the type of this embedded meaning: it can contain either a normal predicate or an embedding predicate, allowing in principle for unlimited recursion and an infinite number of meanings. This recursion is only possible, however, in the second argument position; only objects are allowed as the first argument to a predicate.

$$\text{Meaning} = \begin{cases} (\text{Predicate}_\alpha(\text{Object, Object})) \\ (\text{Predicate}_\beta(\text{Object, Meaning})) \end{cases} \tag{4.2}$$

Hurford (2000)'s semantic model is in a similar vein, based on a simple world of humans and animals, first described by Cann (1993). In Hurford's model, there is further expansion of the types of predicates which can be found, this time not just in terms of the type of patient they take, but also in terms of their valency, or the *number* of arguments they can take. In addition to the dyadic predicates which can be read as transitive verbs, as in Kirby's simulations described above, Hurford also has monadic relationships

[2]It seems that there is a further implicit restriction in the model, which ensures that the same object is never allowed to appear in both places of the predicate. For instance, there are no 'reflexive' meanings like <Agent=Zoltan, Patient=Zoltan, Predicate=Finds>

such as HAPPY and triadic relationships such as GIVE. Recursion is also implemented, but this time through not a whole class of embedding predicates, but just by one special SAY-predicate, which makes the further requirement that its agent must be human[3]. A description of Hurford's semantic model can be seen in 4.3.

(4.3)
$$\text{Individual} = \begin{cases} \text{Human} \\ \text{Animal} \end{cases}$$

$$\text{Meaning} = \begin{cases} (\text{Predicate}_1(\text{Individual})) \\ (\text{Predicate}_2(\text{Individual, Individual})) \\ (\text{Predicate}_{\text{SAY}}(\text{Human, Meaning})) \\ (\text{Predicate}_3(\text{Individual, Individual, Individual})) \end{cases}$$

Batali (2002)'s semantic model differs slightly from those of Kirby and Hurford, in its use of variables, although the representations are clearly still based on predicate logic. Batali's representations are called *formula sets*, and are composed of a predicate and variables, or arguments, as shown in 4.4. Batali distinguishes two kinds of predicates, analogous to those shown in 4.3: monadic predicates, which he calls *properties*, and dyadic predicates, or *relations*. Two formula sets can be combined into another formula set by simply juxtaposing any number of them[4] (represented by the Kleene star notation in 4.4), and further manipulated by altering the mapping of the variables within them, to create more complex meanings. Batali deliberately chooses not to implement recursion directly, but nevertheless the repeated combination of formula sets produces in principle an infinite set of possible meanings.

(4.4)
$$\text{Meaning} = \begin{cases} (\text{Predicate}_1 \quad x) \\ (\text{Predicate}_2 \quad x \quad y) \\ (\text{Meaning})^* \end{cases}$$

Having looked at the predicate logic representations used by Kirby, Hurford and Batali, it is interesting to investigate their semantic models in terms of their semantic content. In particular, what do the predicates and arguments *refer* to, and what *sense-relations* do the meanings have with each other? We might assume that the meaning of the predicates is

[3]All other predicates in Hurford appear to be able to take any individual, either human or animal, as any of their arguments.

[4]Batali has imposed an arbitrary limit of seven formula sets per meaning, for ease of implementation.

that of the common English words which are written identically to them, and that they are used to refer to actions in the model's imaginary world just as the English words are used to refer to actions or objects in the real world. But here we stumble across an important problem which recurs in many of these evolutionary simulations: the agents do *not* use the meanings to refer to actions or objects in their world, because there is no way in the experiment for the agents to access their world[5]. There is, therefore, nothing useful we can say about the reference of the predicates and arguments in these models; they have no denotation at all, because there is no external semantics in the models over which any denotation must be specified.

In order for us to be able to regard a meaning representation as encoding sense relations, at the very least there must be some structure in the representation, so that some relationship, however tenuous, between different elements (meanings) in the representation can exist. There are, therefore, some distinctions made in the models which we could arguably interpret as sense distinctions, particularly the hierarchical division of INDIVIDUAL into ANIMAL and HUMAN in Hurford (2000)'s simulations, which of course finds a parallel in the semantics of many natural languages. Crucially, however, we find that these 'sense' distinctions are *not* available to the agents in Hurford's model, who instead merely have two pre-defined, arbitrary classes of names, one of which can be used as an argument to any predicate, and the other which can be used as an argument to any predicate except SAY.

Overall, therefore, although each experimenter has implemented a structured representation which they have called 'semantic' in these models, there is very little about these representations which relates to sense and reference, and thus very little about them which can be sensibly regarded as in any way semantic, apart from the name itself. Instead, the purpose of the 'semantics' in these models is actually to serve as a blueprint for the syntax, which will then appear to emerge from the simulations. The agents' task is to learn a mapping between representations in two mediums: an existing, unchanging code which the experimenters call semantics, and a new, modifiable, emergent system which they call syntax. As Nehaniv (2000) has pointed out, syntax only develops successfully from unstructured signals in these cases because the signals are coupled with meanings which are already structured, and it is no coincidence that the emergent 'syntactic' structure directly parallels the pre-existing 'semantic' structure.

[5]Indeed, in Batali's model, there is no mention of an external world at all.

	a_0	a_1	a_2	a_3	a_4
b_0	a_0b_0	a_1b_0	a_2b_0	a_3b_0	a_4b_0
b_1	a_0b_1	a_1b_1	a_2b_1	a_3b_1	a_4b_1
b_2	a_0b_2	a_1b_2	a_2b_2	a_3b_2	a_4b_2
b_3	a_0b_3	a_1b_3	a_2b_3	a_3b_3	a_4b_3
b_4	a_0b_4	a_1b_4	a_2b_4	a_3b_4	a_4b_4

Table 4.1: Kirby (2001)'s model of meaning as a two-dimensional matrix with five discrete meanings on each axis.

4.2.2 Vectors and Matrices

The models discussed in this section use a different semantic representation,which is more abstract and less obviously based on a well-known formalism like predicate logic, but yet with meanings that still display a certain amount of the structure necessary for us to discern sense relations between meanings.

Kirby (2001) moves away from explicit predicate logic by introducing meanings which are vectors in two dimensions a and b[6]. Each dimension can range over a specified number of *discrete* values, and so the whole set of meanings, or the meaning space, can be thought of as a matrix, with a finite number of possible meanings, as can be seen in table 4.1, where there are 25 discrete meanings. Kirby (2001)'s model described above is very similar in its representation of meaning to one which was first presented by Steels (1996a). In this model, as in Kirby's, meanings are represented in terms of discrete values of features. Steels explicitly names both the features WEIGHT, SIZE, SHAPE and their respective values[7], but as in previous models, the names are merely mnemonics to help in understanding the model. The only real difference between the meaning space representations is merely that while Kirby's is a two-dimensional matrix with five discrete values on each dimension, Steels' is a three-dimensional matrix with three discrete values on each dimension. Kirby, therefore, has slightly reduced both the dimensionality and the number of possible meanings in the simulations, or cells in the matrix of meaning (25 (5^2) compared to 27 (3^3)), but otherwise the models' meaning representations are identical.

Brighton (2002), in a paper showing how compositional syntax arises under cultural pressures, extends the representations of both Steels (1996a) and Kirby (2001), by creating

[6]The two parts of the meaning could of course still be interpreted as predicate and argument, but this interpretation is no longer built in to the model.

[7]The possible values of the attributes WEIGHT, SIZE, SHAPE are { *oval, round, square* }, { *tall, small, medium* } and { *heavy, light, average* } respectively.

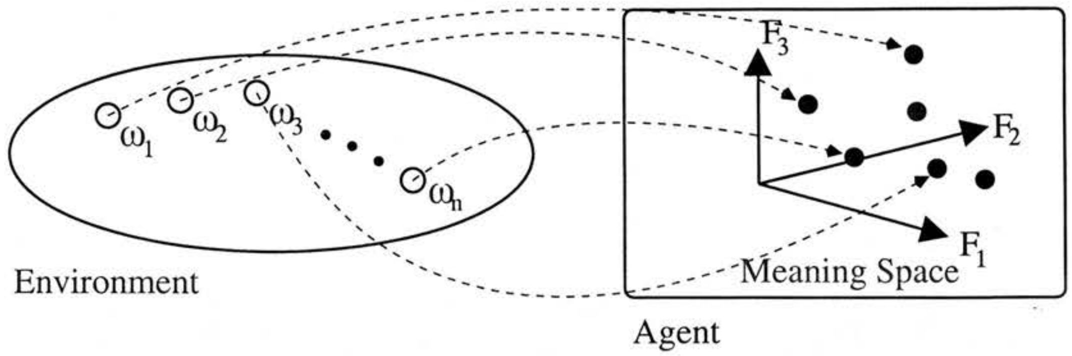


Figure 4.1: The relationship between the situations in the environment and points in the agent's meaning space, from Brighton (2002). The agent's meaning space is constructed as a multi-dimensional matrix, in F dimensions (here $F = 3$), with V discrete values possible on each dimension.

a more general meaning space, which is defined by two parameters: the number of features, or dimensions F, and the number of possible discrete values which can occur on each feature V, as shown in his diagram, which is reproduced as figure 4.1. Both the previous models, therefore, should be considered as specific instantiations of Brighton's more general model of meaning as a multi-dimensional matrix: Steels' can be defined with $F = 3$ and $V = 3$, while Kirby's can be defined with $F = 2$ and $V = 5$. It is important to point out, however, that figure 4.1 is potentially misleading in its depiction of the meaning space; despite its portrayal with apparently continuous axes, the meaning space is indeed constructed as a multi-dimensional matrix, with each dimension $F_1 \ldots F_F$ made up of a fixed, finite number (V) of *discrete* values.

Brighton (2002) also introduces an explicit external environment to the model, which consists of a number of *communicatively relevant situations*. These situations in the environment correspond in turn to distinct points in the discrete, multi-dimensional meaning space, as is portrayed by the dotted lines in figure 4.1. This mapping is specified randomly at the start of the simulation, and never changes thereafter. This representation of meaning as vectors clearly has a different underlying semantic model from the models of Kirby (2000) and Hurford (2000) discussed in the previous section. There is here an explicit external environment, and the meanings therefore appear to have *reference* to objects, or situations in this environment. The meaning space is also explicitly structured, so we can consider relationships between particular meanings, and it might be argued that the meanings do have some kind of *sense*, if we take a rather broad definition of sense as a relationship of any sort between meanings. For instance, meaning $\{1, 2, 2\}$ is related to meaning $\{2, 2, 2\}$ by virtue of the fact that it differs only in the first dimension, being identical in the second and third[8]. On the other hand, neither of these possible sources of semantic representation are quite what they seem; although the meanings appear to have reference, on closer inspection this turns out to be illusory, and although we do find some kind of relation which could be called a sense relation, this is not as great as might at first be envisaged by Brighton's notation.

Moreover, Brighton's generalisation algorithm itself is interesting, because of its great power based on extrapolating from chance correspondences to whole dimensions of meaning space on the basis of feature value identity and difference. We might imagine, for instance, that meaning $\{5, 1, 1\}$ could be considered as 'nearer' to meaning $\{3, 1, 1\}$ in terms of Euclidean distance than it is to meaning $\{8, 1, 1\}$, and therefore that distances relatively near to each other are more likely than distant ones to be considered as the 'same' meaning, but in fact this relationship is surprisingly not used in Brighton's model[9]. If an agent meets two meanings $\{5, 1, 1\}$ and $\{3, 1, 1\}$, both associated with the same signal, it does not use a simple generalisation, marking one signal with both meanings it has met (like $\{[35], 1, 1\}$[10]), nor does it even generalise across a contiguous portion of the meaning space, bounded by the meanings it has met (like $\{[3\text{-}5], 1, 1\}$), but actually it generalises across *all* possible meanings in the dimension where differences occurred ($\{?, 1, 1\}$), as shown in table 4.2. To take a real-world example of features with discrete values, let us imagine that the objects in Brighton's model represent chemical elements, and the first dimension represents the atomic number of the elements. When an agent meets two objects with the same signal, one of which is lithium (atomic number 3) and the other of which is boron (atomic number 5), Brighton's generaliser chooses not to mark the signal with a simple generalisation (*lithium* or *boron*), nor a spatial generalisation including the element which stands between lithium and boron in the periodic table (*lithium* or *beryllium* or *boron*), but generalises dimensionally across all elements, assuming that the atomic number, and the identity of the chemical element, is an irrelevant distinction for this signal. This is a legitimate, if very powerful, generalising strategy,

[8]This much, of course could also be said about the predicate logic representations discussed in the previous section: HAPPY(x) differs from HAPPY(y) only in its argument, as both expressions use the same predicate.

[9] In Brighton, Kirby, and Smith (2003)'s related model, on the other hand, the authors do indeed make use of this distance relationship in the meaning space to derive their measure of compositionality.

[10]The notation I use both in this paragraph and in table 4.2, is taken from the language of regular expressions (Friedl, 2002). In particular, I will make use of the following three expressions:

- $[xy]$ represents either x or y
- $[x\text{–}y]$ means either x or y or any other possible value between x and y
- and ? is a wildcard which matches any one possible value.

Table 4.2: Various methods of generalising over two meanings $\{5, 1, 1\}$ and $\{3, 1, 1\}$

Type	Notation	Members of Generalised Meaning
Simple	$\{[35], 1, 1\}$	$(\{3, 1, 1\}, \{5, 1, 1\})$
Spatial	$\{[3\text{-}5], 1, 1\}$	$(\{3, 1, 1\}, \{4, 1, 1\}, \{5, 1, 1\})$
Dimensional	$\{?, 1, 1\}$	$(\{1, 1, 1\} \ldots \{3, 1, 1\} \ldots \{5, 1, 1\} \ldots \{V, 1, 1\})$

which focuses on the similarities between two meanings and generalises over the differences, but it is important to note that Brighton's agents take account neither of any 'distance' between meanings, nor of how many possible meanings they are generalising over, and therefore that the number of meanings which this merged meaning $\{?, 1, 1\}$ corresponds to, and so the power of the whole generalising algorithm, are both explicitly determined by the particular value of V in each experiment. If V is relatively high, such as in the number of known chemical elements (currently 113)[11], then exposure to just two different values in one dimension causes the agent to assume that *all* different values of that dimension are expressed in the same way.

There are indeed some relationships between the meanings in Brighton (2002)'s model, which might charitably be interpreted as sense relations (although in truth they bear little resemblance to any traditional sense relations such as hyponymy and antonymy), but do these meanings have reference? The environment, and in particular its relationship to the agents' meaning representations, is not as important as it first seems in these models. Although the environment is explicitly linked to the meaning structure, by being defined as the source of the meanings, and represented as such in figure 4.1, on closer inspection we can see that the relationship between environment and meaning actually plays no role at all in the simulations; the agents never interact with the environment in any way, and the environment actually appears to be more of an obfuscatory factor in the model. We have seen in the previous section how the presence of an external environment is necessary for the development of a real semantic system, but now Brighton (2002)'s general model shows us that the *mere* presence of an environment is not enough: it is also necessary for the agents to have some interaction with their environment; without this, there is no way in which the meanings can have reference.

A direct extension of Steels' vector-based method of meaning representation is described by de Jong (2000), whose model is inspired by Cheney and Seyfarth (1990)'s study of vervet monkeys, and consequently whose agents' semantic 'state-action' space has three

[11] The apparent synthesis of element 118 has been retracted by Ninov et al. (2002), its purported discoverers.

state dimensions, representing the presence of a particular predator (S_1), and the agent's horizontal (S_2) and vertical (S_3) positions, together with two action dimensions, corresponding to movements horizontally (A_1) and vertically (A_2). The meanings described by de Jong, which he refers to as *situation concepts*, or patterns in the history of an agent's interaction with its environment, do not fit straightforwardly into Brighton's general model, because his meaning structure is not sufficiently uniform to be defined in simple terms with the two parameters F and V used by Brighton. Although the number of features (F) in de Jong is clearly five, the number of values on each features (V) is not uniform; after all, this model is tailored towards the specific problem of modelling the vervet communication system, rather than the more general problem of meaning creation.

For instance, the predator (S_1) feature has four possible values, representing the three specific predators and the absence of any of them. The other features fall naturally into two pairs, representing the horizontal (S_2 and A_1) and vertical (S_3 and A_2) positions, but each works slightly differently: The vertical positioning feature (S_3) has three possible values, and the vertical action feature (A_2), which defines a new vertical position for the agent, likewise has the same three possible values. The horizontal action feature (A_1), on the other hand, is represented explicitly in terms of movement relative to the current position: either to stay still, or to move one step to the left or to the right, again making three possible values. Because the horizontal action feature does not choose an absolute position, but instead defines its actions relative to the current position, the number of values on the horizontal position feature (S_2) is in principle unlimited[12], although in practice, when the predators appear in the world, they must be sufficiently near to the agents, in terms of their horizontal position, or else the agent's predator sensor does not detect them.

The model described by de Jong, nevertheless, has elements of both sense and reference relations in its meanings. The categories in the state-action space are related to each other using a hierarchical relationship, as we shall see in section 4.2.4, and they are also explicitly grounded in the agents' external world through the extraction of feature values.

4.2.3 Word Webs

Hashimoto (1997, 2001) presents a very different semantic model which is based on sense relationships between words. His focus is on the *sense-making* process and on

[12]In fact, the space used by de Jong (2000) is bounded, but I have been unable to find the limits to the horizontal plane which he used in the experiments.

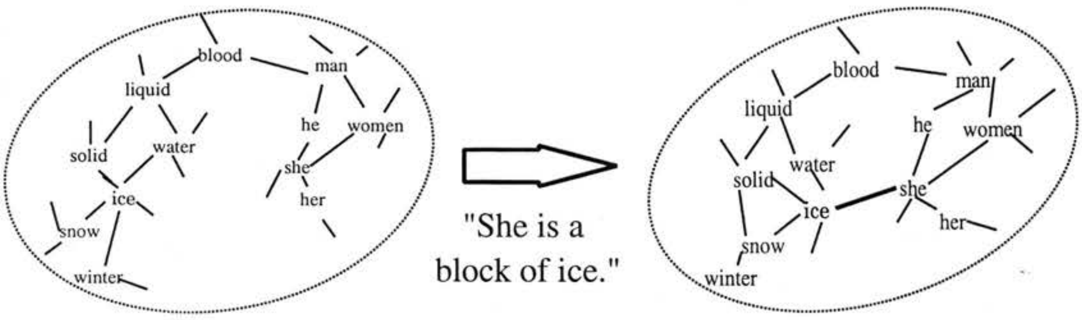


Figure 4.2: Semantic representation as a dynamic word-web, from Hashimoto (2001). The acceptance of the sentence by the agent triggers modification of its word-web.

language as a truly dynamic system, which is modified and remodelled after every communicative or linguistic episode. Interestingly, Hashimoto actually makes no distinction between words and word meanings, which are represented using an enormous word-web, implemented as a dynamic matrix, which models the relationships between words based on patterns of word usage and collocation in particular utterances and in larger texts of utterances, as shown in figure 4.2.

Hashimoto's semantic representation is clearly based on sense relations, although it is worth noting that the only relationship which is actually modelled is an amalgamation of word similarity (a measure of the frequency with which words are used in the same sentence) and word correlation (a measure of the patterns of word appearance in texts); there is again no modelling of even basic hierarchical sense relations, such as those we discussed in chapter 2. As a purely sense-based system, whose relationships are built from word usage patterns, we are not surprised to find that there is no reference at all in Hashimoto (2001)'s model. Again, there is no environment or world outside the agents, so there is no possibility that the words can refer to anything in this external world. This may also hold a clue to the lack of basic semantic notions such as *hyponymy* in this model; there is no way for the agents to discover that the set of referents referred to as CAT (its *extension*) is a subset of those referents referred to as ANIMAL, and in fact there is no way in which such a relationship can be represented in the basic word-web in figure 4.2, without further modifications which could potentially specify the type of the relationship represented by the connections between words.

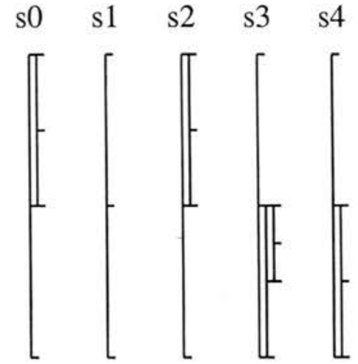


Figure 4.3: Steels' (1996) representation of meaning on discrimination trees. The discrimination trees are built on sensory channels (s0-s4), and are shown with the root of the tree at the left. Each segment of the tree shows the bounds between which it is sensitive.

4.2.4 Trees

In Steels (1996b), however, a different approach is put forward, which has been further developed in Steels (1997, 1999) and Steels and Kaplan (2002), and has been extended by many researchers since, including the work in this thesis: instead of defining a set of meanings which will be used by the agents in their language games, Steels simply defines a framework for representing meaning, on which the agents build their own individual representations. These semantic representations can be represented as a discrimination tree, with each segment showing the bounds between which it is active, as shown in figure 4.3. This semantic representation described by Steels can be clearly seen to have a reasonable number of sense relationships built into it; although not as comprehensive in its inclusion of multiple relationships between meanings as Hashimoto's word-web, Steels (1996b)' meanings have an obvious hierarchical structure, allowing the representation of semantic relationships such as hyponymy (one segment being a subset of another segment higher up the tree). As we saw in chapter 2, a binary tree structure also allows the implicit representation of antonymy, as each segment which has been refined into two subcategories necessarily has two co-hyponyms, which can each be regarded as the other one's antonym.

Steels' model is also closely bound to the environment in which the agents are situated, and, as we shall see in section 4.3, it is actually the main driving force behind the creation of meanings. Each segment on the tree, or category in the semantic representation, is abstract, and yet it also explicitly *refers* to a group of objects in the external world, namely those objects in the world whose feature values fall into the range to which the particular segment is sensitive. The bounds which define each category do not overlap, so the membership of a category is clear and distinct; for each feature value, at each level of the tree, there is only one branch which can be chosen, and only one possible meaning. Of course, because the tree is clearly hierarchical, the same feature value can still have different meanings at different levels of the tree, so a value which falls within the category defined by the upper quarter of a particular tree at the second level will automatically also fall into the category defined by the upper half of the tree at the first level[13], just as in our actual representation of meaning, the subcategory HERON is automatically also part of the larger category BIRD.

The meanings in de Jong (2000)'s model, which I discussed briefly in section 4.2.2, can also be thought of in terms of discrimination trees in a space, although the important difference between de Jong and Steels (1997) is that de Jong's meanings are each defined in all five dimensions of his meaning space at once. By contrast, although the agents in Steels' model do have multiple sensory channels on which discrimination trees are built, the channels are not related to each other multi-dimensionally (although segments on them can be combined to create compound meanings), and meanings are defined in one dimension alone, on each channel individually. Just as Steels' subcategories can be represented as ever smaller one-dimensional lines on the tree in figure 4.3, so de Jong's subcategories can be thought of as ever smaller five-dimensional subspaces on a multi-dimensional tree. Clearly, meanings which are represented as multi-dimensional subspaces are very difficult to represent graphically, and I will not try to do so here, but it is important to note that this multi-dimensionality of meaning has implications for the creation of meaning in de Jong's model, as I will discuss further in section 4.3.2.

We can see, therefore, that the meanings in Steels (1996b)' original model and its subsequent manifestations (Steels, 1997, 1999; Steels & Kaplan, 2002) and modifications (de Jong, 2000) clearly have both sense relations and reference relations, and are therefore the most truly semantic of any of the representations we have seen so far.

4.2.5 Prototypes

Vogt (2000), has implemented a model of meaning which bears some relation to de Jong's model, but with two main differences: it has been physically situated in actual robots, and the categories are the first to be based on the *prototype* model of meaning rather than the classical model, recalling our discussion in chapter 2. Vogt's categories are regions in a four-dimensional meaning space, and a particular meaning is defined by its relative

[13]Every value also falls into the category defined by the root of the tree, but this category is usually ignored, because it is of no practical use in helping the agents make sense of their world, as we shall see in section 4.3.

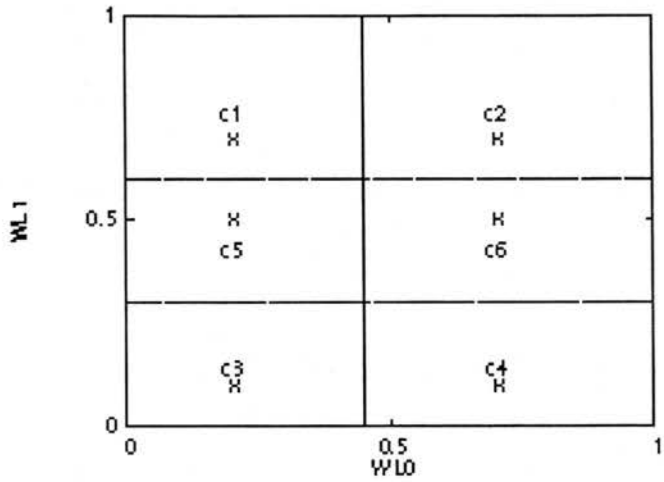


Figure 4.4: A representation of meaning as hyper-rectangles in a four-dimensional space, based on vicinity to prototypes, from Vogt (2000). Only two dimensions of the four-dimensional meaning space are shown, with the location of the prototypes marked by *(x)* and the names of the categories they form by *(c1-c6)*.

vicinity to one of the existing prototype points in the space, as shown in figure 4.4, which depicts just two of the four dimensions in order that the structure of the categories can be easily shown. The regions in Vogt's meaning space always have a hyper-rectangular shape, just as in de Jong (2000)'s model, so there is one important way in which Vogt's model of meaning necessarily deviates from an ideal prototype model; the boundaries between one category and another are clear and distinct, rather than fuzzy, making his model in this respect a compromise between a classical and a prototype representation. There is clearly some sense-like structure in the multi-dimensionality of both de Jong's and Vogt's meaning representations using subspaces, which is perhaps to be expected in structures which are derived explicitly from that of Steels (1996b). We can also see that the meanings represented inside both de Jong's computer agents and Vogt's physical robots have explicit reference to situations and objects which are encountered in the agents' environment.

Finally, I will investigate another different kind of meaning representation, in which meanings are again stored as prototypes. Despite the attractiveness of a prototype theory of meaning in certain situations, very few simulation models actually implement such a model, probably due to the difficulties involved in the representation of the system, and the concomitant processing power needed to run meaningful experiments. Belpaeme

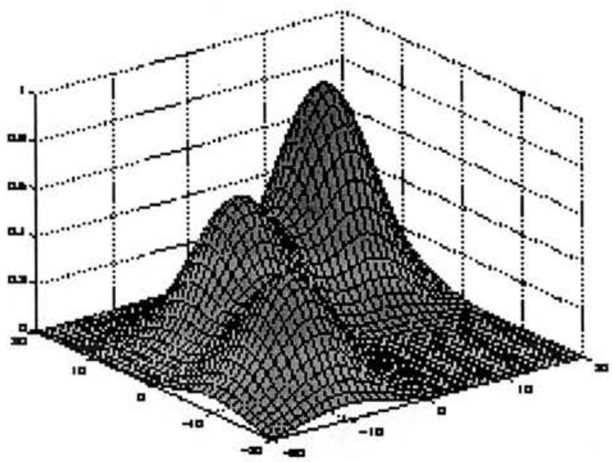


Figure 4.5: An adaptive network in a two-dimensional space, from Belpaeme (2002). This network, which represents one meaning, has three locally tuned units, each defined in terms of their centre, width and height.

(2002), however, in contrast to Vogt (2000), has managed to use a method of representing prototype meanings where the categories have fuzzy boundaries. He uses *adaptive networks* to represent categories, which are based on radial basis function networks (Orr, 1996), as shown in figure 4.5. Each category is represented by a different adaptive network, and each adaptive network is made up of a number of locally tuned units, which define the network. A locally tuned unit is defined by a Gaussian function around its centre; the function is always positive, but its value decreases monotonically as we move away from its centre, producing the characteristic bell-curves we see in figure 4.5. Each unit is also defined by its width, or the steepness of the curve's decline, and its weight, or the value of the function at the centre of the unit; the adaptive network in figure 4.5, which represents *one* meaning, has three locally tuned units with different centres and different weights, although each of the three functions has the same width and so the curves decline at the same rate.

The main advantage in Belpaeme's approach is that the meaning space is not divided into discrete regions, as in all the other approaches in which it makes sense to talk of a 'meaning space'. Instead, we can look at a point in meaning space in terms of the adaptive networks, by interpreting the value given by an adaptive network in response to a stimulus from the environment as a measure of category membership, or as a response to the question "how much of a [category name] is this stimulus?". The value produced by the adaptive network, therefore, naturally provides a fuzzy, graded notion of category membership consistent with that suggested by Rosch (1973). The main disadvantage is that the boundaries between categories barely exist at all, although they could of course be superimposed if required; an adaptive network gives some measure of category membership for every point in the space, so there is no way to say that a particular point is definitely *not-x*, and the boundary between categories cannot be clearly stated.

On the other hand, although Belpaeme (2002)'s model was clearly inspired by, and is primarily focused on, the evolution of colour categories in agents, his simulations actually have little to do with colours in particular and could have been presented as the evolution of any abstract categories; the stimuli received by his agents are essentially just points in a continuous three-dimensional space[14]. His model of meaning has a built-in measure of similarity between meanings based on the weighted sum of minimum distances for all the locally tuned units in each adaptive network, but does not lend itself easily to hierarchical or other sense relations. In common with the simulations we have looked at in the latter half of this chapter, however, the agent's categories are explicitly grounded in their environment, so they can be said to refer to stimuli or objects therein.

4.2.6 Summary

One of the crucial attributes which relates to the expression of reference is the idea that the agents have access to and are able to interact with some kind of external world, and objects therein which can be *referred* to. In the models whose meaning representation is based on predicate logic, such as Kirby (2000, 2002), Hurford (2000) and Batali (2002), this external world is missing, and the semantics presented is merely a code which the agents must decipher. In Steels' (1996a), Kirby (2001)'s and Brighton (2002)'s models, some structure has been added to the meaning representation, which is, as we have seen, a pre-requisite for the implementation of real semantic sense relations; additionally, both Steels and Brighton introduce into their models the notion of an external world, notwithstanding the fact that Brighton's external world is actually more of a distraction than an integral part of his model. Hashimoto (2001) presents a model which explicitly manages without reference, as it builds its semantic structure entirely on word collocations. Although no semantic notions other than collocation can be found, this model clearly has the potential to be extended to encode other semantic relationships reasonably straightforwardly.

By contrast, the meaning structures in Steels (1996b, 1997, 1999), Steels and Kaplan (2002) clearly contain both hierarchical sense relations and a real relationship with an outside world. It is not coincidental that these models are based not on providing an

[14]Belpaeme explicitly defines this space in terms of the L*a*b* space devised by the Commission Internationale de l'Eclairage, where L* represents lightness, a* red-greenness and b* yellow-blueness, but it is not clear that anything is gained by preferring this over a more abstract stimulus space.

innate set of meanings to the agents, but instead on enabling the agents to create their own meanings by providing them simply with a framework for representation and the ability to interact with their environment. Steels' original model has been extended to cover multi-dimensional meaning structures (de Jong, 2000; Vogt, 2000), and more substantially to incorporate representations of meanings as prototypes, with both discrete (Vogt, 2000) and fuzzy (Belpaeme, 2002) boundaries, without losing the properties of reference which are important to a semantic model.

It is clear that these latter, Steelsian, models are the most appropriate on which to build a model of meaning construction, and in chapters 6–9, I present a model based on this which will allow me to investigate how the interpretation of meaning affects the properties of agent-constructed communication systems.

4.3 The Creation of Meanings

In section 4.2, I surveyed many different systems for the representation of meaning in simulations which have been proposed by researchers into language evolution, and looked at how the conceptual systems relate to the Fregean notions of 'sense' and 'reference', which are often used to define meaning. Many types of meaning representation have been put forward, representing both sides of the divide between classical and prototype meanings we encountered in chapter 2, as well as more abstract representations based on predicate logic and mathematics. Having done this, we will now look at those same simulation models, but this time focusing on where the meanings originate and how they are created. Having already discussed the often acrimonious debate between nativists and empiricists, it is perhaps not too surprising to find a parallel, though altogether more amicable, dichotomy in the field of simulations of language evolution, between experimenters who provide a ready-made, 'innate' system of meaning for their agents on one hand, and those whose focus of enquiry is the creation of the meanings by the agents on the other.

In the first category fall the experimental models by Kirby (2000, 2001, 2002), Hurford (2000), Batali (2002), Brighton (2002), Brighton et al. (2003) and Hashimoto (1997, 2001), who all provide some kind of innate meaning representation for the agents at the start of the simulation. We can deal with these models briefly in this chapter, because the creation of meanings does not play a large role, if any, in their simulations. Typically, agents in the first group of models are provided with a finite set of meanings, according to whichever representation of meaning the experimenter has chosen, as I discussed in detail in section 4.2. During the experiments, the agents play two roles, with their exposure to these meanings being slightly different in each: as speakers, they are given a random meaning by the experimenter, which prompts them to produce an appropriate signal; as hearers, they receive both the speaker's signal and the meaning it expressed as a combined signal-meaning pair. The hearer's task is not to produce, but to try to discover the mapping between the two halves of the signal-meaning pair. There is no creation of meanings at all, therefore, unless we include the initial setup of the simulation, when a set of meanings is generated. If at any point new meanings are added to the agents' repertoire, then these too are explicitly generated and given to the agents. We can safely ignore such models, therefore, for the purposes of investigating the creation of meaning by the agents themselves.

On the other side of this particular ideological fence are the models by Steels (1996b, 1997, 1999), Steels and Kaplan (2002), de Jong (2000), Vogt (2000) and Belpaeme (2002); these experimenters provide the agents merely with the *capability* of creating meanings, and investigate the conditions under which they are successful. In these models, the development of the semantic space is an important part of the simulation, and so is much more interesting for our purposes. I will consider each of these models in turn, starting with those created by Steels (1996b, 1997, 1999), who created the basic framework from which all the others have been developed, and I will explore how the agents go about the process of developing their own semantic systems.

4.3.1 Discrimination Games

The basic procedure of agent-based grounded meaning creation, of agents developing meanings based on and relevant to the world they inhabit and the experiences they have, was initially modelled by Steels (1996b), who named it a *discrimination game*, after Wittgenstein (1953)'s famous *language games*. The Steelsian discrimination game is both selectionist, adaptive and minimalist: *selectionist* because the environment in which the game is played, and the dynamics of the game itself apply pressure to the agent's internal representations; *adaptive* because it responds to the results of the game to adapt its own internal representations in various ways; and *minimalist* because the agents in the simulations are provided only with basic operations for meaning creation, and not any intelligent generalisation or language-specific capabilities such as those which have been suggested for human infants and which we surveyed in chapter 3. I will briefly describe the four constituent parts of all discrimination games in the Steelsian paradigm, namely scene-setting, categorisation, discrimination, and adaptation, below, and will then go on to discuss its varying implementation by researchers, who each use slightly different methodologies, just as they used different methods of meaning representation.

scene-setting: the agent is given a specific discrimination task based on its environment, as follows:

- the agent is situated in a world made up of objects or situations, the features of which are in some way detectable by the agent.
- a set of objects or situations, called the context, is presented to the agent.
- one of the objects in the context is chosen to be the *target* of discrimination[15].

categorisation: the agent goes through all the objects in the context, returning for each an association with one or more of its existing semantic representations.

discrimination: the agent tries to find a distinctive category for the target. A category (or a set of categories) is distinctive if it is a valid representation of the target, and is not a valid representation of any other object in the context.

adaptation: the agent modifies its internal conceptual structure in some way; the methods of adaptation available to the agent are typically simple and few.

The processes of scene-setting and of discrimination itself are essentially fixed and identical in all implementations, although the Steelsian abstract model (Steels, 1996b, 1997) has been adapted into the *Talking Heads* experiments (Steels, 1999; Steels & Kaplan, 2002), in which real robots were built which could segment a scene into objects, and extract features from the scene they had developed, rather than being presented with the feature values from the objects. On the other hand, the particular methods of categorisation and adaptation of semantic representations are of course dependent on the particular semantic representation which has been adopted. In the next section, I shall briefly investigate the various implementations of the discrimination game.

4.3.2 Binary Category Splitting

The essential ingredients of the categorisation sub-task of the discrimination game are the reception of feature values from a space of possible values and the translation of these into a new space of possible categories. In Steels (1996b)'s model, and in his subsequent modifications thereof, including the implementation on the Talking Heads robots (Steels, 1997, 1999; Steels & Kaplan, 2002), the agents receive values from a number of different

[15]Steels (1996b) originally named this object the *topic* rather than target, but this term has connotations of conversational units, as well as a linguistic definition as "that element of a sentence which is presented as already existing in the discourse"(Trask, 1993), both of which can be misleading in a purely discriminatory situation.

features. In earlier models, the features were abstract without any specific meanings, but in the robot implementations, the features were pre-defined into spatial and colour characteristics such as AREA, HORIZONTAL and VERTICAL POSITION, HEIGHT, WIDTH, GREYNESS, RGB (the amount of red-, green- and blue-ness in an object), and the number of EDGES and ANGLES in the shape. For our purposes, however, the particular characteristics to which the features correspond are irrelevant, so I will generally regard the features as abstract, unless exemplification with a particular pre-defined characteristic is especially enlightening.

Each feature is independent as far as the agent is concerned, and the values it receives are normalised so that they always lie in the range range [0.0 ... 1.0]. The translation from feature space to category space therefore involves a translation from an infinite number of possible values into a smaller number of categories (although also theoretically infinite). Steels' meaning representation is established on discrimination trees; the agent has a specific sensory channel for each feature, and on each sensory channel can build a separate discrimination tree. Each discrimination tree, therefore, corresponds to a specific feature of the objects in the world, underlying the conceptual independence of the features from each other. Each node on the tree is a category, and corresponds to a particular contiguous segment of the feature value space; the root node of the tree corresponds to the whole of the feature value space, i.e. it is bounded by 0.0 and 1.0 respectively. Categorisation, therefore, is the translation of a continuous feature value into a particular node on a discrimination tree. Of course, a category must exist before it can be used to categorise an object in terms of its feature value, and the agents adapt their semantic representation by one of the following means:

1. a new discrimination tree is created on a sensory channel which has no meaning structure.
2. an existing node on a discrimination tree is chosen, and meanings are added:
    - The region to which the existing node corresponds is split into two discrete segments, equal in size.
    - A new meaning is created for each of the new segments.
3. a node is pruned, or deleted, from the discrimination tree.

Because any created category can potentially be the source of a future refinement, the meanings created through this procedure fall naturally into the hierarchy shown in figure 4.6, which shows a very simple example of a discrimination tree which has been built

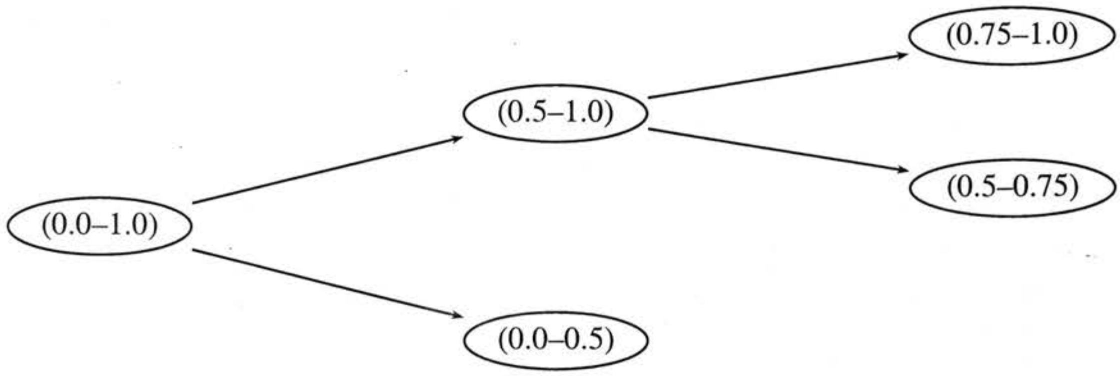


Figure 4.6: The development of categories on an abstract sensory channel shown as a discrimination tree. Each node on the tree shows the bounds between which it is sensitive; the root of the tree is sensitive to the entire feature value range (0.0–1.0); it has two daughter nodes, each of which is sensitive to half of the root node's range. The daughter nodes can also potentially have their own daughter nodes, and so the meanings can easily be represented in tree form.

on an abstract sensory channel. The root of the tree is sensitive to the entire feature value range (0.0 – 1.0); it has two daughter nodes, each of which is sensitive to half of the root node's range. In principle, a category could be divided into more than two segments, and the daughter categories need not have equal-sized sensitivity ranges, but as there is no limit to how fine-grained the distinctions which can be made even with the basic binary category splitting procedure, it seems sensible to stick initially with this framework, which is simple yet powerful, and ideal for exploring the development of meaning in agents. In Steels' models, failure in the discrimination game is the trigger for the adaptation of a sensory channel and the creation of more conceptual structure in the form of more specific categories. There is no pre-definition of which meanings should be created, however; the new categories *may* turn out to be useful in future discrimination games, but there is no guarantee. The agents in this model, therefore, have a mechanism for constructing concepts which are grounded in the environment (Harnad, 1990) and adaptive to their surroundings.

In de Jong (2000)'s models, the agents again receive information on sensory channels, but the feature values are specifically tied to particular representations of the state of the world, the actions of the agent, and an evaluation of the appropriateness of the action, as we discussed in section 4.2 which gives more detail of the structure of de Jong's meaning space. The environment provides a high reward for a specific action in the presence of each of the three different predators, corresponding to the appropriate evasive action taken by the vervets (see table 2.1), and provides high rewards for all actions when there is no predator around. The agents in de Jong's model have a fixed task, which is to create a sufficiently detailed semantic structure to successfully identify the presence of the predators and take the appropriate action.

The categorisation and adaptation processes in de Jong (2000)'s model are very similar to those in Steels (1997), although they may not appear so on first contact. The agents use the same kind of categorisation, in that they check whether or not a particular value falls into a subspace of the overall space, and the categories form a set of discrete categories. Interestingly, however, the original feature values in de Jong's model are not continuous, but already discrete, so it is unclear why he needs to introduce a categoriser which converts continuous variables into discrete ones, except that this kind of categoriser is of course more general and can be more easily used with other problems. The adaptive subspace method, as de Jong calls his process of concept formation, works in the same way as Steels' refinement process, with two main differences.

Firstly, the meanings are created across the whole of the multi-dimensional meaning space; although the task is set up so that the five sensor dimensions provide specific information, the agents do not know about this specificity, and search for general k-dimensional subspaces in whatever space they are provided with. The actual process of concept formation, however, is the same: a hyper-rectangular category is split into two smaller hyper-rectangles of equal size.

Secondly, however, de Jong's agents do not split categories blindly, without regard for whether the new categories will be useful in future, but instead decide whether to split a category or not on the basis of the pre-defined evaluative rewards they receive for particular combinations of states and actions from the simulation itself, and as such they are guided explicitly by a *reinforcement learning* process, which is useful for this kind of fixed problem. An agent is always looking for ways of splitting its meaning structure into subspaces, or *situation concepts*, considering a potential split in each dimension of its state-action space in turn. This potential split, as in Steels' (1997) model, would bisect the particular dimension, resulting into two smaller subspaces, both hyper-rectangles half the size of the original space. The criterion for whether to actually make the split is based on whether there is a significant difference between the distributions of the rewards for the experiences in each of the two potential subspaces. Summarising briefly, once all five dimensions have been investigated, a split takes place, realising the potential subspaces, in the particular dimension for which there is the greatest difference, as long as this difference is above a pre-defined threshold. In this way, the agents adapt their meaning space more quickly to the environment than Steels' agents, who are blindly creating

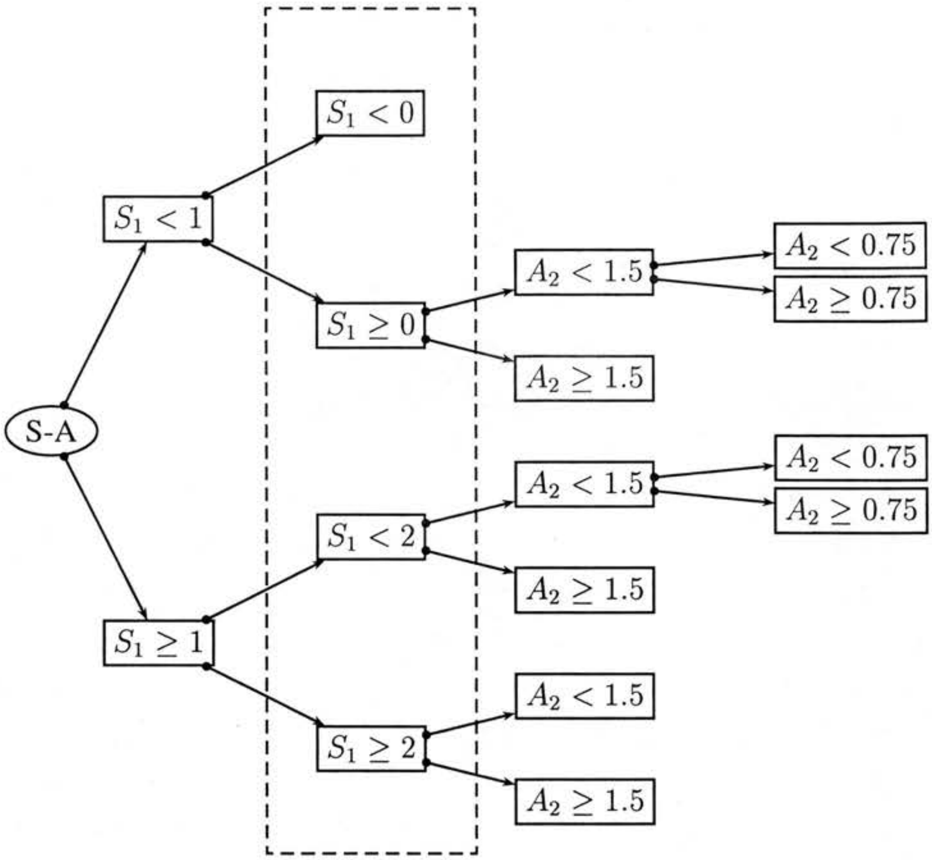


Figure 4.7: De Jong's representation of meaning in a five-dimensional hyperspace using adaptive subspaces. Distinctions have been made in the S_1 and A_2 dimensions.

categories which may or may not be useful to the agents in the future. In contrast, all the new categories created by de Jong's agents are necessarily useful in discriminating experiences; if they were not, then the split which created them would have remained a potential split, and would not have been confirmed.

Figure 4.7 is taken from de Jong (2000), and shows his representation of this meaning creation in five-dimensional space; again, only two of the dimensions are shown to make the figure comprehensible, because the particular task which the agents are set, to respond with appropriate actions in the presence of particular predators, can actually be solved without reference to the other dimensions. Recall that the dimensions of de Jong's meaning representation in a state-action space, as we saw in section 4.2.2, represent information both about the state of the world (S), and about the actions taken by the agents (A). In figure 4.7, the agent has split dimension S_1 into four situations, representing each of the three predators, and the 'safe' situation where no predator is found, which are all shown inside the dotted box in figure 4.7. For each of these situations, except the top one, which represents the safe situation, the agent has also split dimension A_2, which relates to the action it should take in terms of its own vertical dimension, and it has chosen a different action subspace for each of the three predator situations. In other words, figure 4.7 is a (fairly opaque) representation of the fact that the agent has successfully solved the task it was set, that is to find the appropriate action for each of the four different situations it finds itself in.

Although de Jong (2000)'s model solves the problem it was set, and expands Steels' (1997) original model into five dimensions, it remains problematic, on account of its design, which stems from the author's desire to associate the task so closely with the vervet monkey communication system we are now so familiar with. It seems to me that this kind of reinforcement learning paradigm is singularly unsuited to solving tasks of this nature; the agents in the model are not passing on innate knowledge like the vervets, but are being asked to learn from scratch the categories which are important to them, based on the feedback they get from the environment. This is all well and good, but if this was a real world consisting of vervet monkeys and predators, then the feedback they get from the environment would not allow them to solve the problem; rather the first time they chose the wrong action, they would be caught and killed. Reinforcement learning, which is based explicitly on learning from your mistakes, is not particularly useful if the cost of failure is so catastrophic as death, as Oliphant (1999) has noted.

4.3.3 Prototype Manipulation

In this section, I will explore the processes of categorisation and adaptation in the systems which used a prototype model of meaning (Vogt, 2000; Belpaeme, 2002). In both cases, recall that the overall structure of the discrimination game remains broadly the same as that designed by Steels (1996b) (see section 4.3.1), but that their particular implementation of categorisation and semantic adaptation are of course different.

In Vogt (2000)'s model, which is implemented on physical robots rather than inside a computer, categorisation works in the same way as in the models just described; the agents find out the space into which a particular feature vector falls, and return the category which defines this space. It is important to differentiate in Vogt's model between the *prototype*, which is a single point in the meaning space, and the *category*, which is a region in the feature space containing those points which are near the prototype. Although the prototype is the basis of the category, the category itself always has a hyper-rectangular shape, rather than a hyper-spherical one, so that no point in the space falls outside categorisation, and the boundaries between categories are clear and distinct.

Vogt's agents, however, also have a number of different models, or layers, of the same feature space, so that they can have overlapping categories on different dimensions.

Failure in the discrimination of objects is again the trigger for the adaptation of the agent's semantic representation, and this is done by increasing the number of prototypes in the world, up to a maximum resolution of the feature space, which is defined arbitrarily before the experiment. When adding categories, the agent chooses a random feature of the target object which it failed to discriminate, and uses the values of this to double the number of prototypes in the feature space; the new prototypes differ from the existing ones only in their positioning with respect to the chosen feature. For instance, assuming a three-dimensional space, and prototypes already existing at (0.1,0.2,0.3) and (0.9,0.2,0.3), the agent chooses the second feature or dimension, for which the target object had a value of 0.6. Two new prototypes are therefore created at (0,1.0.6,0.3) and (0.9,0.6,0.3), in the same position as the existing ones, except in the second dimension. This has the same effect of splitting the feature space in half, producing categories as hyper-rectangular boxes; the growth of categories can likewise be displayed on a (multi-dimensional) discrimination tree.

In addition, Vogt (2000)'s agents also update their prototypes when they succeed in using a category during a communicative episode, which I will discuss further in chapter 6. This addition to the model means that the categories adapt not only to failure, but also to success; the categories are adapted by shifting them slowly towards the feature vector which was successfully used. In order for communicative success to trigger the adaptation of categories, the agents must receive *feedback* from the model which allows it to evaluate communication. As we have already seen, the existence of feedback to language learners is hotly disputed (Bowerman, 1988) and is therefore absent from the model of agent-constructed communication which I will present in chapters 6–9. In effect, then, the agents in Vogt's model are using an instance-based learning technique (Mitchell, 1996), creating new prototypes when discrimination fails, and supplementing this with reinforcement of successful meanings when communication succeeds.

Belpaeme (2002)'s model of categorisation is different from all the others we have seen, because it is based on fuzzy prototypes; this means that for any object or stimulus, a measure of categorisation is returned. Category membership is no longer a binary yes/no decision, but is a matter of degree. When playing the discrimination game, therefore, Belpaeme's agents choose not the category which matches an object, but the category which *best* matches the object. Thereafter, the procedure is similar to that which we have seen before; if the target object's category is different to the category of all the other objects in the context, then the game succeeds.

Failure triggers the adaptation of the network in one of the following ways:

1. If the agent has no categories, then one is created which describes the topic, consisting of a network with one locally tuned unit, centered on the sensory representation of the topic.

2. An existing category is adapted to better represent the topic.

3. A new category is created.

A pre-defined threshold value inside the agents determines whether a new category is created or whether an existing one is modified. Adaptation of a network uses much the same procedure as the creation of a new network: a new locally tuned unit is added which is centered on the topic's representation. Belpaeme also tunes existing categories when they are successful, so that they are more like the topic, and has the locally tuned units decay over time, so that unused units eventually drop out of the category definition.

4.4 Summary

In this chapter, I have reviewed many recent simulations of the evolution of language, paying particular attention to their models of meaning representation and meaning creation. All of these models claim to have a 'semantic' meaning space, yet on closer inspection, the majority of the models had categories which were innate, pre-specified by the experimenters themselves, and had no reference to any external world at all. Many of them, in addition, had no sense relations of even the most basic type; there were no relationships at all between one category and another, which were instead atomic, individual items. The meanings in these models are not under the control of the agents at all; only the experimenters themselves can create new meanings or delete obsolete ones, and the meanings can only appear or disappear from an agent's repertoire by 'magic'. In summary, the semantic models of many language evolution simulations are simply not semantic at all, but are instead merely a rudimentary coding system, which the agents in the experiments use as a template with which to decode items expressed in another medium, namely the signals.

On the other hand, there are a sizeable number of experimenters who have made an effort to incorporate some kind of realistic semantic systems, by including an external world of objects which the categories refer to, and by providing various different methods for creating meaning based on the agents' experience in this world. Meaning creation in these models is based on the *discrimination game*, in which an agent's task is to return a category which describes one particular object, distinguishing it from another set of objects. Experimenters have used both classical categorisation and prototype categorisation successfully for their semantic representations; both have advantages and disadvantages, but for the remainder of this thesis I will use my model of semantic representation and creation, which is explored in more detail in the next chapter.

CHAPTER 5

Discrimination Trees

" [Meanings] originate in sensory categories, and are grounded in the iconic and categorical representations that make it possible for you to pick out those sensory categories." (Harnad, 1996, p.41)

5.1 Introduction

In this chapter, I move on to exploring in detail the basic model of semantic representation and creation which is used in the experiments in this thesis. In section 5.2, I look at the creation of meaning through representing the features in an external environment, how meaning is grounded through these representations and how its creation is driven by the failure to discriminate objects in their environment from each other. I investigate the properties of the model in detail, showing how the process of meaning creation works, and the conditions under which it fails. In section 5.3, I show how this simple method of grounded meaning creation, which is very successful at picking out objects from each other, gives rise to independent and divergent semantic structures between agents, even those who inhabit the same environment. I then propose measures to describe the differences in the resultant conceptual structues, which will be used extensively in the experiments described in chapters 7 – 9.

Although one of the underlying principles of this thesis is to assume as little as possible in the model, it seems inevitable, in order to keep the basic interaction process simple, that the agents are able to consider individual objects as separate items, and moreover they have a disposition only to consider full objects when interacting with their environment, rather than parts of objects. This assumption is of course very similar to Macnamara

(1982)'s *whole object bias* which I discussed in section 3.3.2, although it is not confined to the domain of word learning.

5.2 Investigating the Properties of Discrimination Trees

Nearly all the models of independent, grounded meaning creation which have been presented in the literature are based on the initial model described by Steels (1996b), as we have seen above. Steels' initial model is abstract enough to be reasonably simple to implement, yet also produces a semantic model which bears some resemblance to my own, and allows the investigation of semantic properties and processes which are not possible in innate, pre-defined systems, as we saw in section 4.3.

I have already briefly described the workings of the generic Steelsian discrimination game in section 4.3.1, but a more detailed look at my particular model is in order here, as the concepts used are important in understanding how the agent's categorisation works. The discrimination game works in the same way as already described, with agents attempting to distinguish one particular target object from a larger context of other objects, using the same subtasks of *categorisation*, *discrimination* and *adaptation* as in section 4.3.1. The initial environment consists of just one agent, and a population of twenty objects, which the agents are able to recognise as individual items. Each object is defined with a fixed number of characteristics or features; the features themselves are completely abstract, and although it is possible to conceptualise them as things which 'make sense' to humans such as colour or size, they are in fact represented in the model as real numbers between 0.0 and 1.0, which are pseudo-randomly generated with a uniform distribution.

Categorisation

The agent first categorises the objects, by translating the feature value of each object into a category using the discrimination tree on its sensory channel, to find the leaf node λ within whose range the feature value falls. Categories, or *meanings* will be given henceforth using the notation [*sc-path*], where *sc* identifies the number of the sensory channel, and *path* traces the path from the tree root to the node in question. Each node on a discrimination tree is either a leaf (terminal) node or it has two sub-branches; in $path$, 0 signifies that the lower of these branches is traversed, and 1 signifies that the upper of these branches is traversed. Figure 5.1, for instance, shows a simple discrimination tree on sensory channel 0, with not only the range within which each node is sensitive, but also the meaning using to which this corresponds in the above notation.

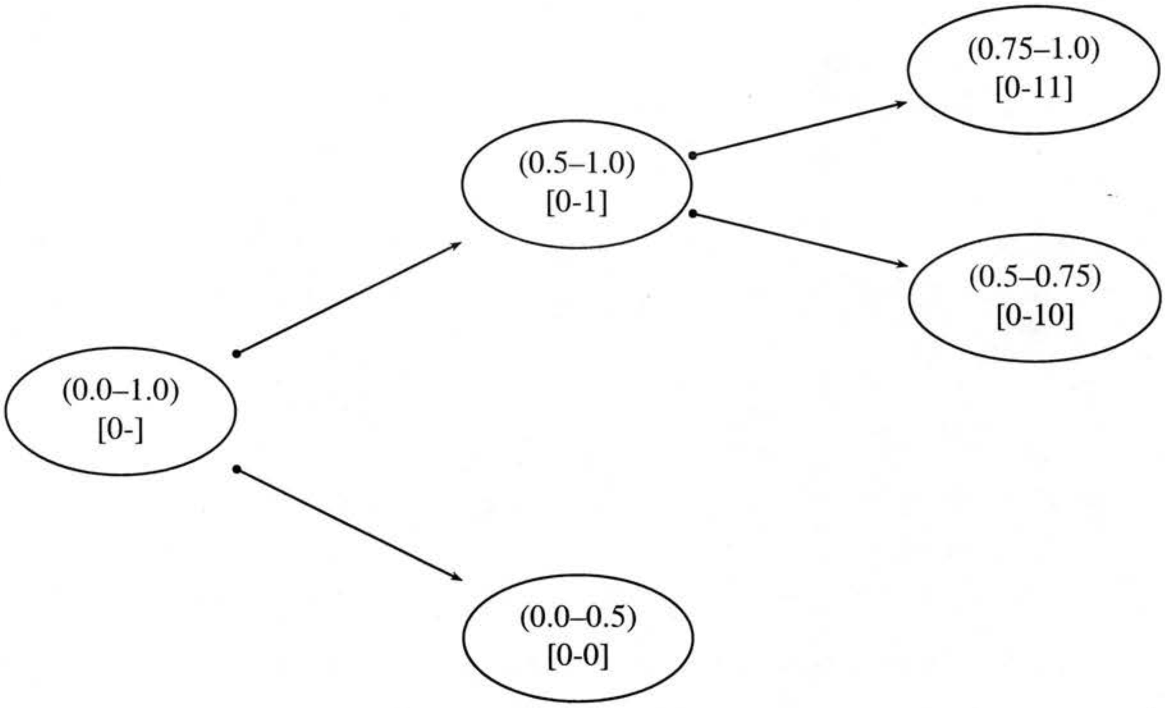


Figure 5.1: A simple discrimination tree, which is being built on sensory channel 0. Each node shows not only the bounds between which it is sensitive, but also the meaning to which it corresponds.

Object	Feature Value	Meaning
A	0.43276	0-0
B	0.87423	0-11
C	0.12098	0-0
D	0.50012	0-10
E	0.31419	0-0

Table 5.1: An agent categorises objects as part of the discrimination game, using the discrimination tree on the sensory channel shown in figure 5.1 to translate the feature values into categories.

Table 5.1, for example, demonstrates the result of translating the feature values shown for objects A–E into meanings on the discrimination tree shown in figure 5.1; the leaf node whose range contains object A's feature value 0.43276, for example, is the lowest node on the tree, which is reached from the root by traversing the lower of the two branches, hence the category $0 - 0$.

Discrimination

After categorising all the objects, the agent investigates the meanings with which it has described the objects, in order to try to find a *distinctive category*, which is defined as follows:

distinctive category: a category which is valid representation of the target object, and is not a valid representation of any other object in the context.

If a distinctive category is found, then the discrimination game succeeds. Only one distinctive category is needed to distinguish the target from the context; if there is more than one sensory channel, and more than one distinctive category exists, then the agent chooses one of them at random to act as the distinctive category in this game. If no distinctive category is found, the the game fails, and the agent adapts its conceptual structure in response to this failure, Table 5.1, for instance, shows an agent categorising objects as part of a discrimination game, using the discrimination tree shown in figure 5.1. Two example discrimination games based on these categorisation might proceed as follows:

1. if the target object in this game was B, which should be discriminated from the context $ACDE$, then the game will succeed, with $0-11$ as the distinctive category;
2. if the target object was C, which should be distinguished from the context $ABDE$, then the game would fail, as the category which describes A, $0-0$, is not distinctive, because it does not distinguish C from A or from E.

Adaptation

As we saw earlier, a sensory channel is refined by splitting the leaf node which categorises the target object, λ, into two further discrete segments. In game 2 above, therefore, the agent would adapt its semantic representation by refining the node which categorised C, namely $0-0$, and creates two new subcategories, $0-00$ and $0-01$. This procedure can of course happen repeatedly, and can in principle provide an unlimited number of categories whose range is ever smaller. Although the discrimination tree is a very simple mechanism of meaning creation, it is very powerful, and is ideal for the abstract representation of real hierarchical semantic structures such as those shown in figure 2.1.

5.2.1 A Basic Model of Discrimination

In this section, I will go through the working of my meaning creation simulations in more detail; the environment is set up in a very simple fashion initially, but this will be extended and developed as we progress through the remainder of this chapter. Each of the objects has just a single feature, and the agent has one corresponding sensory channel; this means of course that there is effectively no channel choice when a discrimination game fails. I assume initially that the size of the context is fixed at the minimum of two objects, i.e. the target must be distinguished from just one other object in the model. The simulations are run for an arbitrary 300 discrimination games, and at regular intervals throughout the following measures are taken:

discriminative success (δ) the percentage of successful games;

unique discriminability (ψ) the percentage of objects in the model which can be distinguished from *all* the other objects in the world.

Figure 5.2 shows twenty different simulations for the model plotted on top of one another, with the cumulative *discrimination success* rate shown with solid lines, and the *unique discriminability* of the model shown with dashed lines. We can see that in all the simulations the percentage of successful games rises rapidly from zero to 80% in about 20–30 games, before raising further towards 100% more slowly, but that the unique discriminability of the model is much more variable, still varying between 50–70% after 300 games.

5.2.2 Discriminative Success

If we take a closer look at the dynamics of these simulations, we can see how the agent is evolving the semantic representation on its sensory channel to take account of the actual feature values of the objects in the world. At the start of the simulation, the sensory channel is unrefined, and so will categorise all objects as $0-$, with no path component to the meaning. This necessarily leads to the failure of the first discrimination game and

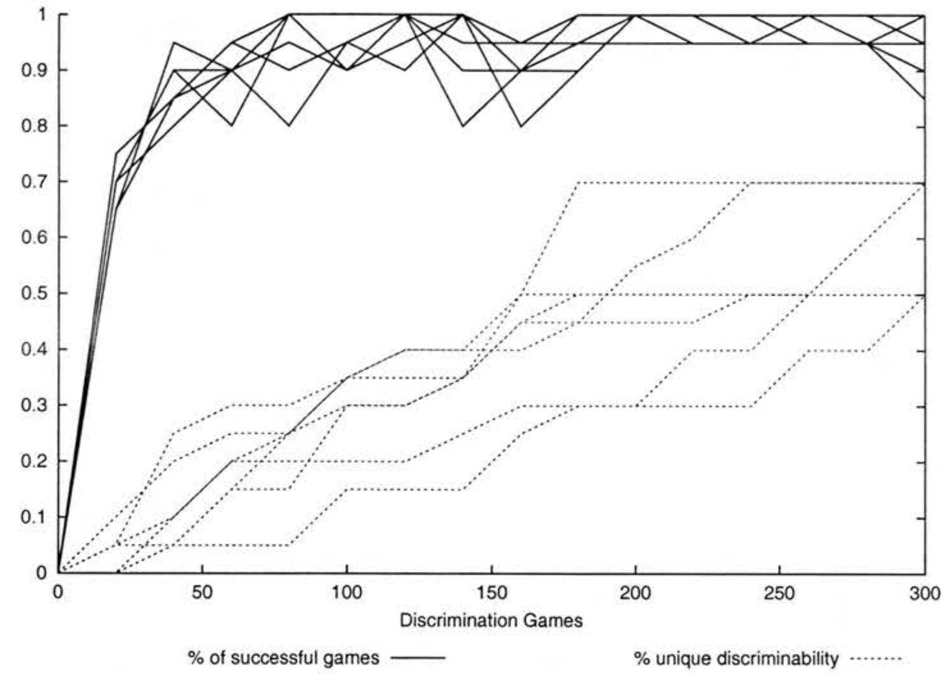


Figure 5.2: Discrimination success δ and unique discriminability ψ. The world contains one agent and 20 objects defined with one feature. The context size is fixed at two objects.

thus also to the refinement of the channel and the creation of two new meanings $0-0$ and $0-1$, with the ranges (0.0–0.5) and (0.5–1.0) respectively. A second discrimination game then takes place, and, given the uniform distribution of the objects' feature values, we could assume that it is likely that about ten objects (or half the total number) have feature values below 0.5, and about ten have feature values above 0.5[1]. Two objects are then chosen at random to be the context, and one of them is then chosen to be the target. In order for the game to succeed, the objects need to be in different categories, and the probability of this happening is:

$$\left(\frac{1}{2} \times \frac{1}{2}\right) \times 2 = 0.5. \tag{5.1}$$

The refinement of the channel has therefore already increased the chance of success from zero to 50%. If the second discrimination game is a success, then no further refinement takes place, and the probability of success in the third game remains at 50%. But of course we can't assume that we will always get successes at odds of 50–50; it would be

[1]Increasing the number of objects will clearly improve the likelihood that the distribution actually does approximate to 50–50.

akin to getting 'heads' every time we tossed a coin. Instead, we investigate what happens when the discrimination game fails again. Failure is triggered by both objects having the same category, and so we must again refine the sensory channel. The agent will refine whichever node categorises the target, so splitting either:

- the lower node $0-0$, whose range is (0.0–0.5), into $0-00$ and $0-01$ (0.0–0.25 and 0.25–0.5 respectively);
- or the upper node $0-1$, whose range is (0.5–1.0), into $0-10$ and $0-11$ (0.5–0.75 and 0.75–1.0 respectively).

Nothing rides on this choice, so I will assume the second option, so that after two discrimination game failures, we have a sensory channel like that already shown in figure 5.1.

The failure of a discrimination game always triggers the refinement of a channel, which in turn always changes the probability of success for the following game. If we continue with our worked example, with the channel as in figure 5.1, we have three possible leaf categories for the target ($0-11$, $0-10$ and $0-0$), with respective probabilities of 0.25, 0.25 and 0.5. The probability that both target and context will be in different categories in this game:

$$\left(\frac{1}{4}\times\frac{3}{4}\right)+\left(\frac{1}{4}\times\frac{3}{4}\right)+\left(\frac{1}{2}\times\frac{1}{2}\right)=0.625. \tag{5.2}$$

Although the development of the sensory channel has been a fairly straightforward progression up to now, it changes after the next failure. Remember that we have a channel like that shown in figure 5.1, with three terminating categories. If we now have a failure, the increase in probability of the next game being a success is dependent on how the sensory channel is refined. This in turn is dependent on the category of the target: if the target falls into the larger category (0-0), then the tree is refined into a symmetrical tree with a depth of two levels, and the probability increases to:

$$\left(\frac{1}{4}\times\frac{3}{4}\right)\times 4=0.75, \tag{5.3}$$

but if the target falls into either of the smaller categories ($0-10$ or $0-11$), and this node is then refined, the probability will be:

$$\left(\frac{1}{8}\times\frac{7}{8}\right)+\left(\frac{1}{8}\times\frac{7}{8}\right)+\left(\frac{1}{4}\times\frac{3}{4}\right)+\left(\frac{1}{2}\times\frac{1}{2}\right)=0.65625. \tag{5.4}$$

We can see immediately that every discrimination game failure, and corresponding refinement of the sensory channel, continues to increase the probability of the following discrimination game being a success, although the absolute increase in probability is dependent on how the sensory channel itself has been refined; I investigate these and similar properties concerning the growth of discrimination trees in more detail in chapter 7. In other words, the agent 'learns' only from its mistakes, and not from its successes. Armed with this information, it is not hard to see from this how the percentage of successful games rises so rapidly (see figure 5.2), and so constantly across simulations: refinements always take place where the target is categorised, and this is inevitably going to reflect the distribution of the feature values of the objects in the world.

5.2.3 Unique Discriminability

In contrast, however, the unique discriminability measure is much more volatile. Unique discriminability, denoted in this thesis as ψ, is defined as the percentage of objects which can be distinguished from *all* other objects in the world by an agent. If an object is uniquely discriminated, then a description of it is effectively the same as identifying it in the world — there is one object alone in the world to which a particular category corresponds.

Returning to our example, after one failed discrimination game, the sensory channel is refined into two segments, and the chances of a success in the next game has risen to 50%. However, there are likely to be about ten objects in each segment, which still cannot be differentiated from one another. Even after the tree has been refined a number of times, we do not get any discriminability at all until one of the categories in the tree categorises only one object. The *maximum* amount of discriminability $max(\psi)$ at any time is given by the following equation:

$$max(\psi)=\frac{n}{o}, \tag{5.5}$$

where n is the number of nodes (categories) on the discrimination tree built on the sensory channel, and o is the number of objects in the world.

We can see that n is initially equal to one, and is incremented after each discrimination game failure. The probability of subsequent failure, however, *decreases* after every actual failed game, and so it can take some time to achieve even potential unique discriminability. The actual unique discriminability measure ψ is always going to be smaller than $max(\psi)$, because although each refinement of a sensory channel increases n and therefore $max(\psi)$, it does not necessarily make any difference to ψ itself. In the simplest case, if a channel has not been refined at all, yet the value of every object in the world is less than 0.5, then the initial refinement will create an extra category, but this will not increase discriminability at all. The likelihood of these redundant categories[2] increases as each tree is refined more and grows deeper. Essentially, the variation in unique discriminability in a model with just one feature reflects the distribution of the feature values; if they are evenly spread around the segmentation points of the bisecting channel, then the level of unique discriminability will be relatively high, but if they are clustered together, it would take longer for an agent to develop categories to a sufficient depth to discriminate. In the limit, in a model with just one feature, the level of ψ will eventually climb to 100%, but only after the number of categories on the discrimination tree exceeds the number of objects in the world. In real human languages, however, it is very rare to have any unique discriminability at all, as words do not pick out *individuals* in the world, but rather *kinds* or *properties* (Lyons, 1977), with the possible exception of proper names, which might form a counter-example (Hurford, 1999, 2001, 2003).

5.2.4 Multiple Sensory Channels

In the model just described, the objects in the world were defined with just one feature, and the agents had just one corresponding sensory channel. In this section, we introduce more features, and describe how the dynamics of the model change to reflect this. Figure 5.3 shows a simulation of a model where each object is defined with five features.

Although the percentage of successful games is similar to that in figure 5.2, the additional sensory channels have resulted in the discriminability of the model being much reduced, staying at zero in nearly all the runs, and with one object uniquely discriminable ($\psi = 5\%$) in a couple of simulations. After the first game in a simulation has, of necessity, been a failure, one of the channels is chosen at random to be refined, leaving

[2]The redundancy of the categories is only in terms of unique discriminability; the new categories might well enable future discrimination games to succeed.

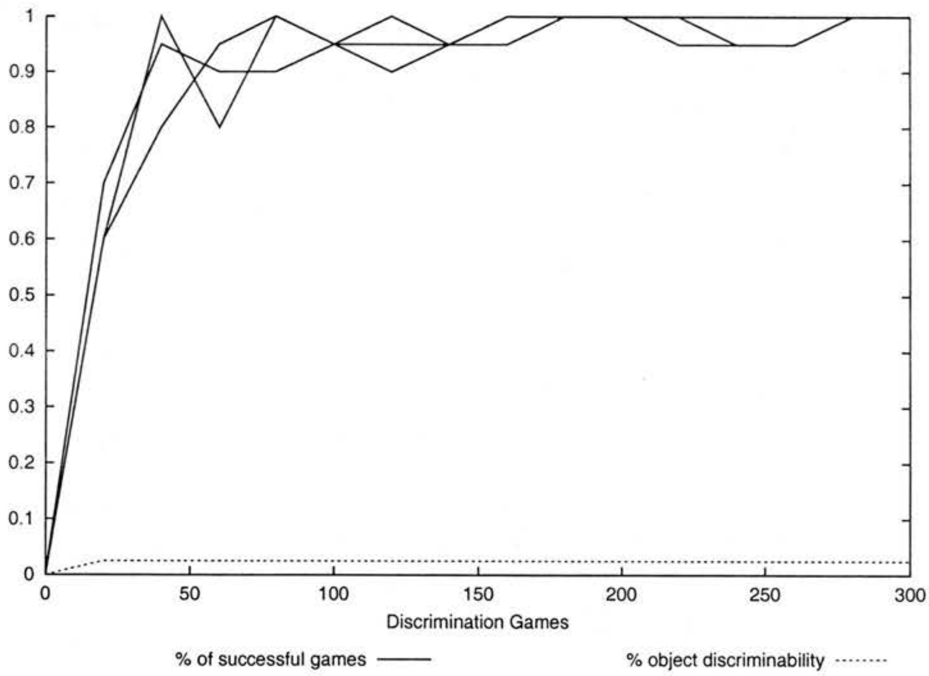


Figure 5.3: Discrimination success δ and unique discriminability ψ. The world contains one agent and 20 objects defined with five features. The context size is fixed at two objects.

the agent now with four unrefined channels, and one channel with two categories. In effect, the situation is the same as in the previous simulation, as only this refined channel can be used for categorisation, and so the chances of success are still 50%. After the next failure, however, we do not get an automatic refinement of the same channel, but one from the five is chosen at random again[3]. The agent has an 80% ($\frac{4}{5}$) chance of refining an unrefined channel, and then having three unrefined channels, and two channels which each have two categories, and a 20% ($\frac{1}{5}$) chance of refining the channels on which meanings have already been created, then having four unrefined channels, and one like the channel shown in figure 5.1. We have already looked at refining one channel, so let us assume now that we have a situation with two refined channels. The probability of the target object and context object being in different categories is now increased because *either* channel (or indeed both) can be used to distinguish the objects:

(5.6)
$$\overbrace{\frac{1}{2} \times \frac{1}{2}}^{\text{Both}} + \overbrace{\left(\frac{1}{2} \times \frac{1}{2}\right) \times 2}^{\text{Either}} = 0.75$$

[3]In the basic model, there is no bias on the choice of channel, so each is equally likely.

After only two failed discrimination games, we can see that the probability of success in the following game is already 75% if two channels have been refined. With one channel, reaching this level of likelihood occurs only after at least three refinements, and then only if the refinements resulted in a symmetrical tree. The increased number of features, therefore, has increased the chances for succeeding in discrimination games, resulting in even higher rates of discriminative success in figure 5.3 compared to those in figure 5.2. If each channel is refined just once, as in the above example, then the probability of success in the following discrimination game is

$$1 - \frac{1}{2^c}, \tag{5.7}$$

where c is the number of sensory channels. Clearly, this probability rises to near-certainty very quickly (with just five sensory channels, it is already 96.9%). However, as we saw in the previous section, success in the discrimination games is not unrelated to the overall unique discriminability of the simulation. More successes, and therefore fewer failures, mean that the agent does not have to 'learn' very much, and does not refine its sensory channels. These channels have just two categories each, and even on the rare occasions that these discrimination games do fail, it is unlikely that they will develop sufficiently deeply so that one of the categories would contain just one object.

The agent will, therefore, struggle to achieve any sort of unique discriminability at all, as we see in figure 5.3. In a couple of simulations, the agent has managed to be able to discriminate one object from all the others after 250 games, but none at all in most simulations. In fact in most longer simulations, ψ remains at 0% even after 1000 games, and with no prospect of increasing, because the discrimination success rate itself is at, or very near to, 100%. Increasing the number of features, then, improves the success rate of the discrimination games themselves, but the very lack of failures in the games means that the agent hardly needs to develop its sensory channels, and so the absolute unique discriminability is correspondingly very low.

Unique discriminability, the ability to identify objects from *all* other objects in the world, is *not* therefore necessary for successful discrimination games in this model, where the objects need only be discriminated from a subset of the objects in the world. Given that unique discriminability is very rare in human language, a low level of ψ in the simulations is welcome with respect to the relative realism in the semantic structures which are constructed by the agents.

5.2.5 Larger Contexts

Instead of increasing the number of features available, we could have modified the basic model by tinkering with the actual discrimination games themselves, specifically by altering the size of the context from which the target should be differentiated. Further simulations have been run, still using a model of one agent and twenty objects with one feature/sensory channel, but this time increasing the context from two (the target and the context) to five, ten and even all twenty objects.

The changes that an increase to a context of five objects gives are immediately obvious, if we look at the probabilities of the discrimination game succeeding after the initial inevitable failure. The agent has just the one sensory channel, refined once into two segments, and the probability of the next game succeeding is:

$$2 \times \left[\frac{1}{2} \times \left(\frac{1}{2}\right)^4\right] = 0.0625, \tag{5.8}$$

as there are only two possibilities for success, the target occurring in one category and all four other objects occurring in the other category (and vice versa). In contrast to the situation in section 5.2.4 when the objects were described with five features and we quickly reached a near-certainty of discriminative success, increasing the context produces instead a near-certainty of discriminative failure. The flip-side of this high probability of initial failure, of course, is that the agent has to refine its sensory channel, and so 'learns' more. After a second refinement, the probability of the next game succeeding has increased to:

$$\left[\frac{1}{2} \times \left(\frac{1}{2}\right)^4\right] + \left(2 \times \left[\frac{1}{4} \times \left(\frac{3}{4}\right)^4\right]\right) = 0.221, \tag{5.9}$$

which is already a large increase on the previous game, although still much lower than in the previous simulations with a minimal context size. The low probability of success means that the agent will continue to refine its sensory channel, leading to more categories on the channel, and therefore a much greater likelihood that unique discriminability will result.

In general, the probability of the next discrimination game Δ succeeding on a particular sensory channel c, $P_\delta(c)$ can be expressed as:

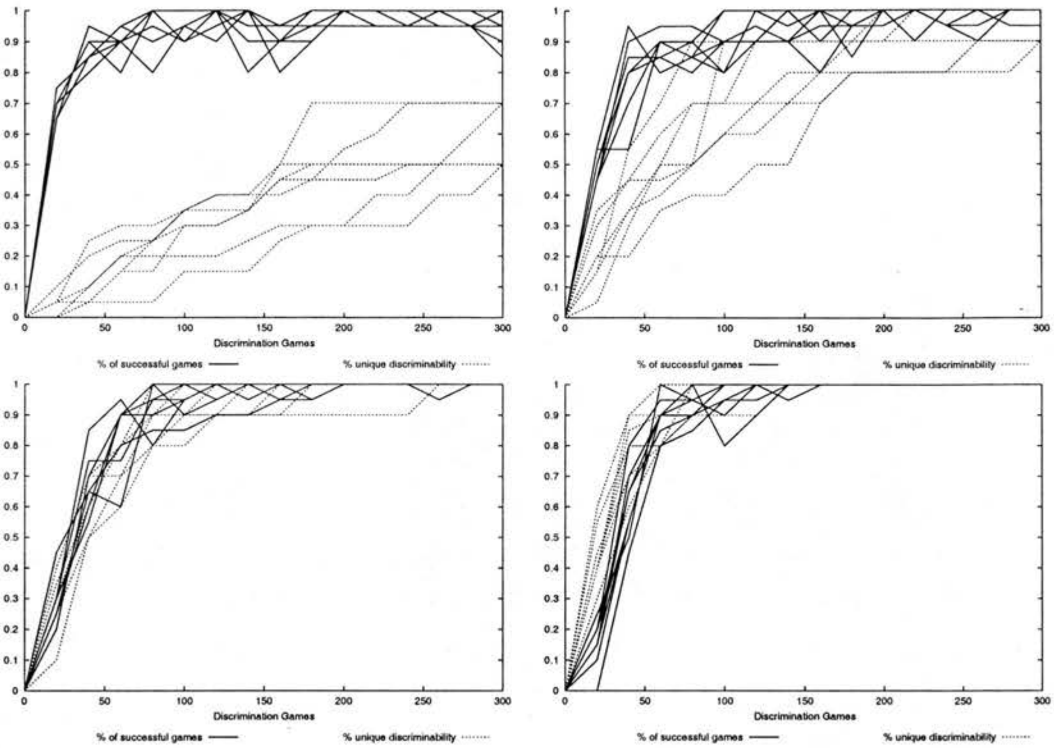


Figure 5.4: Discrimination success δ and unique discriminability ψ. The world contains one agent and 20 objects defined with one feature. Reading from left to right and top to bottom, the context size is fixed at 2, 5, 10 and 20 objects respectively.

$$P_{\Delta}(c) = \sum_{i=1}^{n} p(i) \times (1 - p(i))^{z-1}, \tag{5.10}$$

where n is the number of categories on the discrimination tree, z is the size of the context, and $p(i)$ is the probability of the target occurring in category i, which is equivalent to the *range*[4] of category i.

Figure 5.4 shows the results of these simulations for contexts with two, five, ten and twenty objects, the last of these representing every object in the population. All of these simulations are still run with just one feature for each object, and therefore one sensory channel for each agent. We can see clearly from figure 5.4 that increasing the number of objects in the context reduces the success rate in the games early in this simulation, but this in turn stimulates refinement of the sensory channels and therefore higher unique discriminability.

[4]The difference between the upper and lower bounds of the category.

However, we also need to look at the number of possible discrimination games which can exist in a particular simulation. For any given target object, this is equivalent to the binomial coefficient $\binom{n}{r}$. In our simulation, if o is the number of objects in the population, and z the size of the context, then, taking account of the target object, this is equivalent to $\binom{o-1}{z-1}$, giving the number of different possible discrimination games in a simulation as:

$$o \frac{(o-1)!}{(z-1)!\,(o-z)!} \tag{5.11}$$

Note that if the size of the context z is the whole population, then there are only o possible discrimination games, and each one is in effect an exercise in testing the particular target object's unique discriminability. In these extreme circumstances, it is perhaps not surprising to see that the unique discriminability rates are extremely high.

5.2.6 Multiple Channels and Larger Contexts

As we have seen, increasing the number of features and increasing the size of the context tend to have opposite effects on the results of the simulations, so it is interesting to investigate the effects of both phenomena applied at once. This time we have run simulations with five features, and again looked at the results for different context sizes. To investigate the probability of success in the discrimination games, we need again to take account of the fact that a discrimination by any or all of the features is sufficient, so we can make use of equation 5.10, which gives the probability for a particular sensory channel being able to discriminate, to derive the probability P_Δ that the discrimination game Δ will succeed, or the probability that any of the channels (i.e. not none of them) is able to discriminate, as follows:

$$P_\Delta = 1 - \prod_{j=1}^{c} 1 - P_\Delta(j), \tag{5.12}$$

where c is the number of sensory channels possessed by the agent, and $P_\Delta(j)$ is defined above in equation 5.10. Figure 5.5 shows that, as we might expect, unique discriminability is much reduced, except when the context is very large, when each discrimination game is essentially testing the unique discriminability of the target object.

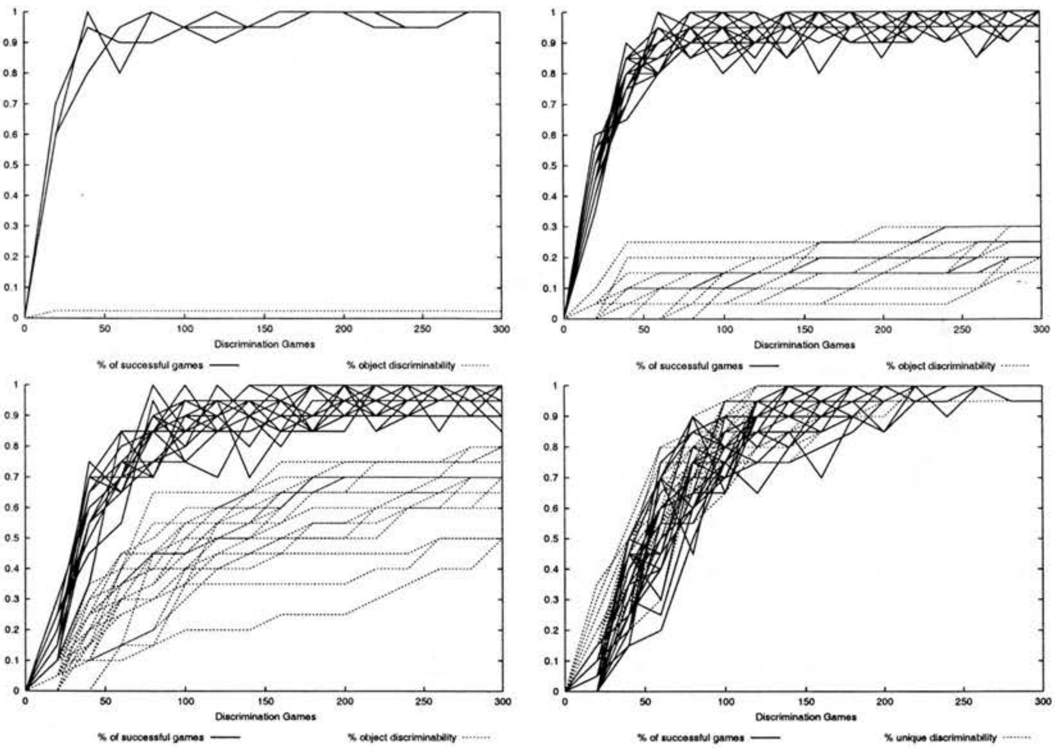


Figure 5.5: Discrimination success δ and unique discriminability ψ. The world contains one agent and 20 objects defined with five features. Reading from left to right and top to bottom, the context size is fixed at 2, 5, 10 and 20 objects respectively.

We can see from these results that discrimination trees are a very successful and efficient method of achieving unguided meaning creation based on observation of the world. After a small number of episodes, we invariably find that the discrimination success rate approaches 100%. The parameters we have chosen to vary, i.e. the number of features which define the objects and the size of the context, affect the speed at which success is achieved, and the extent to which *particular* objects are identified from all others in all possible contexts, which we have called *unique discriminability*.

As we increase the number of features, discrimination success is achieved more quickly, and without any unique discriminability of objects. As we increase the size of the context, we naturally make the discrimination task more difficult, and so success is achieved more slowly, and with higher rates of unique discriminability. If we increase both factors, we find that with small contexts, the effect of the increased number of features is higher, so success occurs quickly and without much discriminability, but as the context size increases, its effects override the feature effects, and we find slower success and higher discriminability. The agents have a mechanism for constructing concepts which is grounded in the environment, is based on their experience, creates meanings which are useful to the agents in allowing them to discriminate between the objects they find.

5.3 Divergent Semantic Structures

Although the trees are developed by the agents in response to their interactions with the environment, there is explicitly no determinism in how they grow. If we create a world in which objects are defined by five features, and then expose two agents to it, we find that their individual meaning representations, as shown in the discrimination trees, can be very different from each other, as we can see in figure 5.6. Both agents have created different meaning structures, based on the same number of discrimination games in the same world. We can see straight away, for instance, that agent one has developed the first three channels to a greater extent than the second agent, who in turn has developed the fourth and fifth channels more extensively. It is helpful to quantify the amount of similarity of two agents' meaning structure, and we do this by averaging the similarity of the discrimination trees built on each of their sensory channels. In greater detail, if $k(t, u)$ is the number of nodes which trees t and u have in common, and $n(t)$ is the total number of nodes or categories on tree t, then we describe the similarity between any two trees t and u using the following formula:

$$\tau(t, u) = \frac{2k(t, u)}{n(t) + n(u)} \tag{5.13}$$

The tree similarity $\tau(t, u)$, therefore, is the proportion of all the nodes on the trees which are shared by both trees. Note that equation 5.13 uses a slightly different measure of tree similarity τ from that described in A. Smith (2003a), shown in equation 5.14, in which, for each tree separately, the proportion of its nodes which were also on the other tree was calculated, with the tree similarity being the average of the two.

$$\tau'(t, u) = \frac{1}{2}\left(\frac{k(t, u)}{n(t)} + \frac{k(t, u)}{n(u)}\right) \tag{5.14}$$

In the vast majority of cases, both these equations τ and τ' produce similar results, although τ is always slightly higher than τ', but there is an important difference when we are comparing one tree which is vastly more refined than the other. The denominator term $n(t)$, the number of nodes on a tree t, is equivalent to $2r$, where r is the number of refinements which has taken place. Crucially, however, the first refinement of any tree is always the same, as it splits the range [0.0 ... 1.0] into [0.0 ... 0.5] and [0.5 ... 1.0]. Therefore, if we compare a tree t which has been refined once, to another tree u which has been refined x times ($x >= 1$), it is necessarily true that $n(t) = 2$, and that $k(t, u) = 2$, in all cases. In equation 5.14, which compares each tree separately, the first term inside the bracket is therefore always equal to 1, which therefore means that the value of τ', in turn, will necessarily be greater than 0.5, even if the other term is very small.

Using the equation in 5.13, in effect widens the distribution of the possible values of τ, because the level of meaning similarity depends much more significantly on the number of nodes on the second tree u as well. For instance, in the same example as above, equation 5.13 reduces to $4/(2 + 2x)$, and we can see that if x is high, then τ is very low, as is appropriate, and is not restricted by an artificial lower bound of 0.5. In cases in which the trees being compared are not grossly dissimilar in the scale of their foliage, however, τ and τ' produce very similar values for the level of tree similarity.

I then use the general measure of tree similarity τ in 5.13 to develop an overall measure of *meaning similarity* σ between two agents, by averaging over all their sensory channels. If a_{ij} identifies the discrimination tree on channel number j for agent i, and each agent has c sensory channels, then the meaning similarity $\sigma(a_1, a_2)$ between agents a_1 and a_2 is defined as follows:

$$\sigma(a_1, a_2) = \frac{1}{c}\left(\sum_{j=0}^{c-1} \tau(a_1j, a_2j)\right) \tag{5.15}$$

If two agents a_1 and a_2 have identical conceptual structures, where $\sigma(a_1, a_2) = 1$, then we say that their meanings are *synchronised*. It is important to note that both agents whose meaning representations are shown in figure 5.6 are successful in the discrimination games, and so their representations are equally good descriptions of their world, although their mutual meaning similarity σ is only 68%. This model of concept creation, then, satisfies one of our main goals, namely that the agents are not given innate meanings, but can create inventories of basic concepts individually, based on their own experiences.

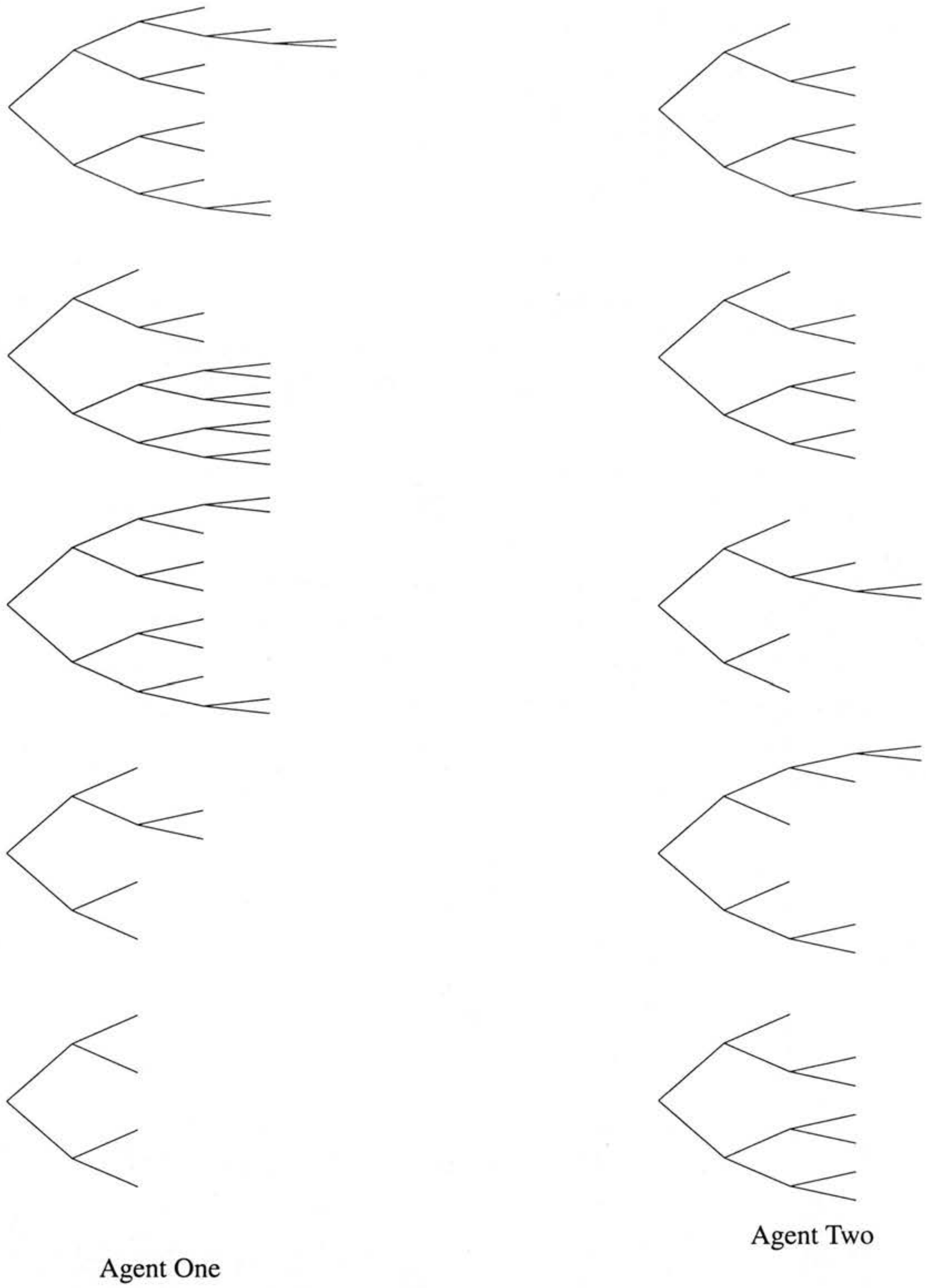


Figure 5.6: Two agents each have five sensory channels, with which they construct different representations of the same world.

5.4 Summary

In this chapter, I have described how meanings are built through the binary splitting of sensory channels, and are structured on these sensory channels in dendritic structures called discrimination trees. We have seen that this simple method of meaning creation is very effective at allowing individuals to ground their meanings in their own experience, to create relevant meanings which allow them to discriminate between objects in their world. High levels of discrimination success δ and very low levels (often zero) of unique discriminability ψ are achived with more features; as the context size increases, δ rises more slowly, and higher rates of ψ are found.

Importantly, there is no pre-definition of their meanings, so each agent creates its own semantic representation individually; many different representations are equally good at discrimination games, and so are equally good descriptions of their world. Finally, I have presented the measures τ and σ, which represent the similarity between two discrimination trees (τ) and between two agents' semantic structures as a whole (σ); both of these will be made further use of in the experiments I go to describe in subsequent chapters.

CHAPTER 6

Communication

"Successful communication does not depend, then, on the communicator and addressee having exactly the same representation of the utterance ... " (Origgi & Sperber, 2000, p. 167)

6.1 Introduction

In chapter 4, I investigated a number of different models in the evolutionary linguistics literature which attempt to grapple with the problems of representing meaning in virtual agents, and of creating and modifying semantic structures under external pressures of some sort. We saw that many of these methods were in fact semantic in name only, but that others, notably those which include an external world which the agents can interact with and which their meanings can refer to, did indeed produce structures which can justifiably claim to be truly semantic both in name and in nature.

I then went on to introduce a model of meaning creation which is driven by the process of discriminating objects in the world from each other, and which produces a hierarchical, dendritic, semantic structure through the repeated binary splitting of categories into sub-categories. In this chapter, I will continue to develop the agent-based model introduced in chapter 4, to explore communication between agents, and in particular I will address the following questions.

1. What does it mean to say that a communicative episode is successful?
2. How do agents choose a word to represent a particular meaning?
3. Conversely, how do they decide how to interpret a word they hear?

4. Can agents achieve communicative success without explicit meaning transfer?

In section 6.2, I discuss the communicative episode in the abstract, looking at the components which any model of communication must contain, and show that one of the most crucial characteristics is a division between public and private knowledge. If there is no such division, and more particularly if one agent's semantic representations are either transferable or visible to others, then the whole model is subject to the *signal redundancy paradox*. In section 6.3, I go on to explore how we can decide whether a particular communicative episode has succeeded or not, and show that we must use some indirect measure based on reference identity. I then discuss some of the implications of this kind of evaluation, particularly in terms of the actual learning of words by children which I looked at in chapter 3, and address important issues like the provision of joint attention and corrective feedback to learners. In section 6.4, I describe previous work on the evolution of communication and vocabulary, which show the importance of lexical bidirectionality and of accommodating the hearer in a communication system. In section 6.5, I go on to describe the lexicon at the heart of my model, which takes account of the findings described in section 6.4, and of the need to avoid the signal redundancy paradox; the strategy is based on Oliphant and Batali (1997)'s obverter, but because I allow the agents access only to their own minds, I call the strategy *introspective obverter* .

6.2 The Constituents of Communication

Communication is the exchange of a message, containing some sort of 'information' between two parties. In using this working definition of communication, I am deliberately avoiding any mention of how the message is exchanged, what format it takes, or what roles the two parties play, though clearly these need to be fleshed out in any detailed description. Importantly, there are two different roles in a communicative episode: the instigator of communication, and the recipient. For communication to take place, there must be both an instigator, and at least one recipient; if either is missing, then there is no communication. No matter how loudly and often the shipwreck survivor on a desert island cries, if nobody is there to hear him then he is not communicating. In the rest of this thesis, I will refer to the communicative parties as *speaker* and *hearer*, and although clearly this usage gives possibly undue pre-eminence to the role of spoken language as a means of communication, rather than sign language or other systems, I would contend that this is certainly justified in linguistic terms, as speech is undoubtedly the primary mechanism through which languages have developed and evolved over generations of human activity.

Hauser (1996) gives an informative overview of how authors from different disciplines have approached the problem of defining communication, and although he assumes the distinction between speaker/instigator and hearer/recipient without further comment, he goes on to suggest that the concepts of *information* and *signal* are central to most of the definitions he gives. We can tentatively adopt Hauser's summary, therefore, and with the addition of the participants of communication, we can preliminarily conclude that the four fundamental constituents of communication are as follows: a speaker, a hearer, a signal, and some information. A communicative episode consists of the speaker producing a signal, which is in turn received by the hearer, and some information, or meaning, the exact nature of which is under further debate. The signal carries some sort of informational content, but this cannot be the source of all the information in the message. We shall see shortly that it is important that the meaning of a message is derived from the context in which it is uttered, as well as from the properties of the signal used to express it, but first let us explore these constituents of communication in more detail, to check whether they are needed in a model of communication, and whether these constituents are sufficient to delimit a model of communication.

6.2.1 Signals

The evolution of signals has received much attention in the literature, not only in terms of the evolution of the innate alarm calls such as the (in)famous vervet monkeys we have already discussed in section 2.4.2, but also in terms of the specific evolution of the main human communicative channel of speech, and thereafter the evolution of the particular speech sounds which are used in human languages. Lieberman (1984), indeed, argues that the development of the specifically human vocal tract, characterised, amongst other features, by a low larynx, a long jaw, a large and rounded tongue, and multiple resonating cavities (oral, pharyngeal and nasal), was the driving force behind the emergence of modern humans. Lieberman argues that these characteristics of human speech provided adaptive benefits in terms of more successful communication, and particularly in terms of the *production* of distinct sounds.

More recently, within the framework of computational simulation which is used in this thesis, de Boer (2001, 2002) investigates the evolution of *particular* sounds as communicative signals; he shows how the dynamics of a population of language users necessarily leads to the creation, over time, of a system of vowels which is optimal in communication by each vowel being almost maximally distinguishable from each other. Figure 6.1 shows an example of a vowel system with just three vowels, this number being widely regarded as the minimum number of vowels in a human language (Katamba, 1989;

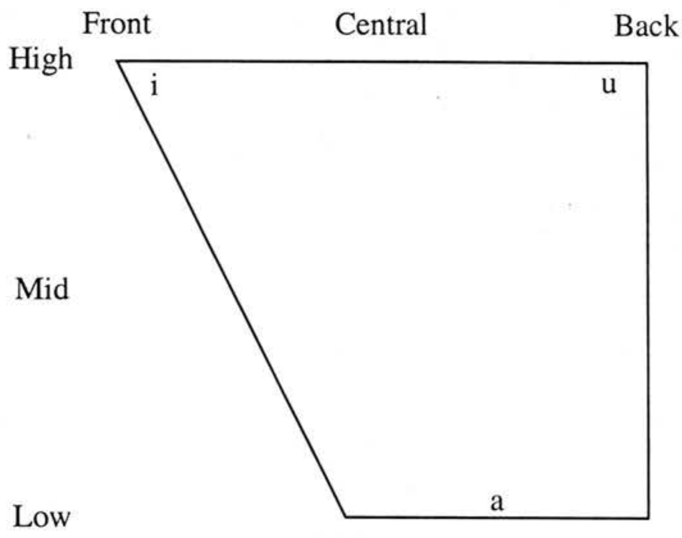


Figure 6.1: A basic system of three vowels which are maximally distinct from each other, as seen in languages as diverse as Aleut and Quechua.

Crothers, 1978)[1], and which is found in languages as diverse as Aleut and Quechua, amongst others. The vowel systems which emerge from de Boer's simulations show a marked similarity with these actual systems, even when there is no optimisation of vowel difference built in to the simulation, strongly suggesting that self-organisation is an important feature of the structure of human sound systems.

In this thesis, however, I am concentrating particularly on investigating the relationship between meaning and communication, rather than the evolution of communicative signals themselves. I assume, therefore, the prior existence of a simple set of signals, which are made up of random combinations of the lower case Roman letters [a-z]. Each signal, or utterance, must be arbitrarily at least two letters long, though a signal can in principle be any length. The letters are merely symbols in an alphabet of salient features which are used to create signals; although they could be thought of in language terms as representing phonemes, they could equally well be considered in terms of other features of communication such as eyebrow raising, nostril flaring, eye opening and ear retraction, which occur, for example, in the signals of rhesus monkeys (Hauser, 1996) .

Importantly, I am also assuming that the task of classification of signals based on their similarity and difference is error-free and automatic; agents can express and receive signals without any error, so that the form in which the speaker utters the signal is exactly

[1]There are however, exceptions to this rule: Vaux and Pəsiypa (1997), for instance, describe the Cºəžə (Tswydzhy) dialect of Abkhaz, which has only two vowel phonemes /ɑ/ and /ə/, and even then the phonemic status of ə is disputed, leaving Cºəžə as a possibly univocalic language. It does, however, have 58 different consonant phonemes to compensate for the paucity of vowels!

the same as that in which the hearer receives it. It is of course a trivial exercise to add noise to this procedure, but I avoid doing so because I am interested in the *inherent* uncertainty of meaning in the fundamental communicative model, and wish to avoid the distractions of this additional uncertainty. The concepts of similarity and identity are absolutely fundamental to all communication systems and central to their analysis, as is widely recognised in psychology (see, for example, Tversky (1977)) and elsewhere. If agents cannot reliably decide whether one signal is to be regarded as the same as another, there is no possibility of generalising across the different situations in which the signal is heard, and no way a meaning can emerge. On the other hand, if agents cannot reliably decide that two signals are different, and every signal is essentially the same as every other one, then likewise meaningful communication is impossible. For this reason, endowing the agents with the ability to recognise whether two signals are identical or different seems a reasonable, and indeed essential, minimal step to take, which allows us to concentrate the exposition on the role of meaning in communication.

6.2.2 Meaning

Given that communicative signals carry some sort of information from speaker to hearer, we must consider what kind of information this is likely to be. The obvious answer is that the signal has a meaning, and it is this meaning which the speaker intends to transfer to the hearer when uttering the signal. Indeed, it would not be unreasonable to assume that the reason for communication is the desire to transfer meanings between conspecifics, via the medium of the signal. There is a widespread assumption amongst researchers using computational modelling to investigating questions of the evolution of language (see Nowak, Plotkin, and Jansen (2000), Hurford (2000), Batali (2002), Brighton (2002), K. Smith (2002b)), nicely exemplified by the following quotation from Kirby (2002), namely that a linguistic utterance consists of the explicit conjunction of a signal and a meaning.

> "The utterances that the individuals produce and learn from in these simulations are pairs of strings of letters and meaning representations." (Kirby, 2002, p. 176)

In this section, I shall explain why I think making this tempting assumption is unrealistic and moreover leads to an unwelcome paradox about the nature of communication and its constituent parts.

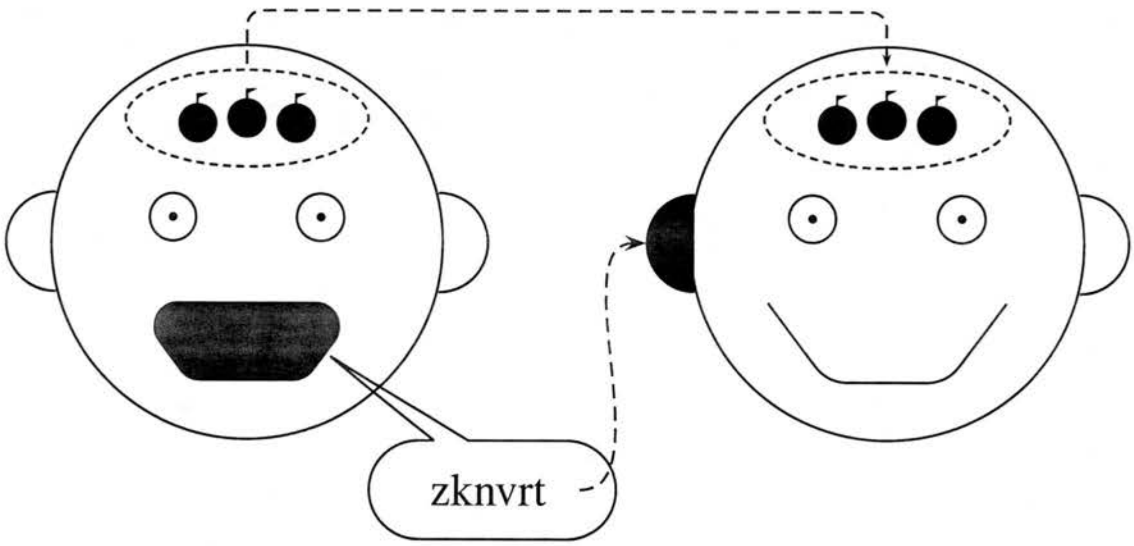


Figure 6.2: A communicative episode which consists of the transfer of a signal *"zknvrt"* and a meaning THREE APPLES from speaker to hearer.

Explicit Meaning Transfer and the Signal Redundancy Paradox

Figure 6.2 shows a schematic diagram of the linguistic transfer in such a communicative model, where the utterances are "pairs of strings of letters and meaning representations" Kirby (2002, p.176). We can see that the speaker (on the left of figure 6.2) utters a signal "*zknvrt*", which is received intact by the hearer, as described above. Simultaneously, the meaning in the speaker's brain, represented in figure 6.2 by three symbols meant to resemble apples, is transferred directly to the hearer's brain. Because the utterance is structured, consisting of two explicitly linked parts, it is a trivial task for the hearer to learn the association between these two parts, between the signal "*zknvrt*" and the meaning THREE APPLES. In chapter 3, we explored the problem of how children learn the meaning of words when they are learning a language, particularly in relation to the Quinean indeterminacy of translation, which shows that there are always an infinite number of meanings which could logically be consistent with the information received, and in relation to the many different mechanisms which have been proposed in the psycholinguistic literature to explain how children solve this paradox effortlessly. If we compare the idealisation of communication shown in figure 6.2 to the discussions of chapter 3, we can see clearly that there are a number of stark and troubling problems.

Firstly, if the meanings are explicitly and accurately transferable by an unspecified telepathic medium, as shown in figure 6.2, then it is clear that the signals such as "*zknvrt*" are not being used to convey the meanings. But what, then, is the experimental role of the signals in such a model? More bluntly, what does the presence of signals add to the model? I hope it is clear, in fact, that the inclusion of such signals is a complicating factor, which adds nothing at all to the communicative process; no information is being transferred between agents by the signals that could not equally well be transferred through the meanings alone. There seems little reason to posit the existence of a system of signals which serves only to replicate another system of meanings, and so, given that the signals are redundant, it seems reasonable to assume that we can remove them from the model completely; after all, the agents can still transfer meanings between each other, and there is now no need for them to expend time and energy worrying about learning a redundant additional system of signalling. The signal-less model, therefore, now consists of a speaker and a hearer, communicating to each other through the telepathic transfer of thoughts (meanings) between their respective brains. But here we stumble upon another serious problem: having removed the redundant signals, the communicative aspect of the model now bears very little resemblance to the communication system we are trying to simulate, namely human language, in which signals play an unarguably crucial role. This problem, which I call the *signal redundancy paradox*, can be summarised as follows:

signal redundancy paradox If the meanings are transferable, then the signals are redundant; but if the signals are removed, then the system no longer represents a (human-like) communication system.

Fortunately, there is a straightforward way out of the signal redundancy paradox, once we realise that it is based on a false premise. The whole problem of signal redundancy only exists if we assume that meanings are transferable; if they are not, then the information cannot be passed directly from one brain to another, but must instead be encoded into a signal. This, of course, is exactly what happens in actual communication systems, where the communication process consists not of the transfer of meaning and (redundant) signal, but only of the transfer of the signal. The exact meaning, as we have seen in chapter 3, must be derived by the hearer from somewhere else, prompted of course by the signal, and by additional factors such as the situation or context in which the signal is heard.

So how do the agents know which meaning to associate with a particular signal? Having established that meanings cannot be transferred, we must conclude that the agents (and by extension, children learning words while acquiring language) infer them from elsewhere. The most obvious, and most general source for this is the environment in which the agent lives; as this is already our source for the construction of meanings, as we saw in chapter 4, using the environment neatly reinforces our model of meanings being grounded, not just in terms of their creation, but also in terms of their communication (Harnad, 1990). It is worth noting, however, that the need to infer meanings from some external source like the agents' environment has interesting implications for the experiments like those described by Kirby (2002). As we saw in chapter 4, these models contain no environment or indeed anything which could be considered both accessible and external to the agents, so the 'meanings' cannot be inferred from elsewhere, and must necessarily be abstract, pre-defined tokens. Because they have no reference, they cannot be communicated except by explicit meaning transfer. Avoiding explicit meaning transfer, therefore, implies that the agents must have access to an external world which they can experience. More than this, however, it implies that there must be at least three levels of representation in the model, as shown in figure 6.3 and described below:

- an external environment, which is public and accessible to all;
- a private, agent-specific internal representation of meaning;
- a set of signals, which can be transmitted between agents and is in principle public[2].

The external environment, from which the agents' experiences are derived, provides the motivation and source for the creation of meanings which will allow the agents to distinguish between particular situations in the environment. Meaning creation itself, however, is a *private, agent-specific* process, based on the particular experiences an agent has. The meanings in an agent's mind do map to the situations in the environment, but this mapping is not perceptible to others; agents cannot read each other's minds directly, but only have access to them indirectly, through their communication process, just as human beings do through language. Communication involves the creation of public signals, which map to the private meanings in an agent's mind; I assume that the signals can be transferred and received without error, while recognising that this assumption is a simplification. It is important that the internal, agent-specific semantic representations are kept private and invisible to others. Moreover, it follows that the mappings between public and private, i.e. between both situations in the environment and meanings on the one hand, and between meanings and signals on the other, must likewise be inviolable, and for this reason both mappings are shown below the dividing line between private and public across the centre of figure 6.3.

If the private sections of the model become public, then the model unfortunately reduces to the equivalent of one which contains the signal redundancy paradox described above.

[2]Of course, each communicative episode does not necessarily involve a broadcast to *all* the agents.

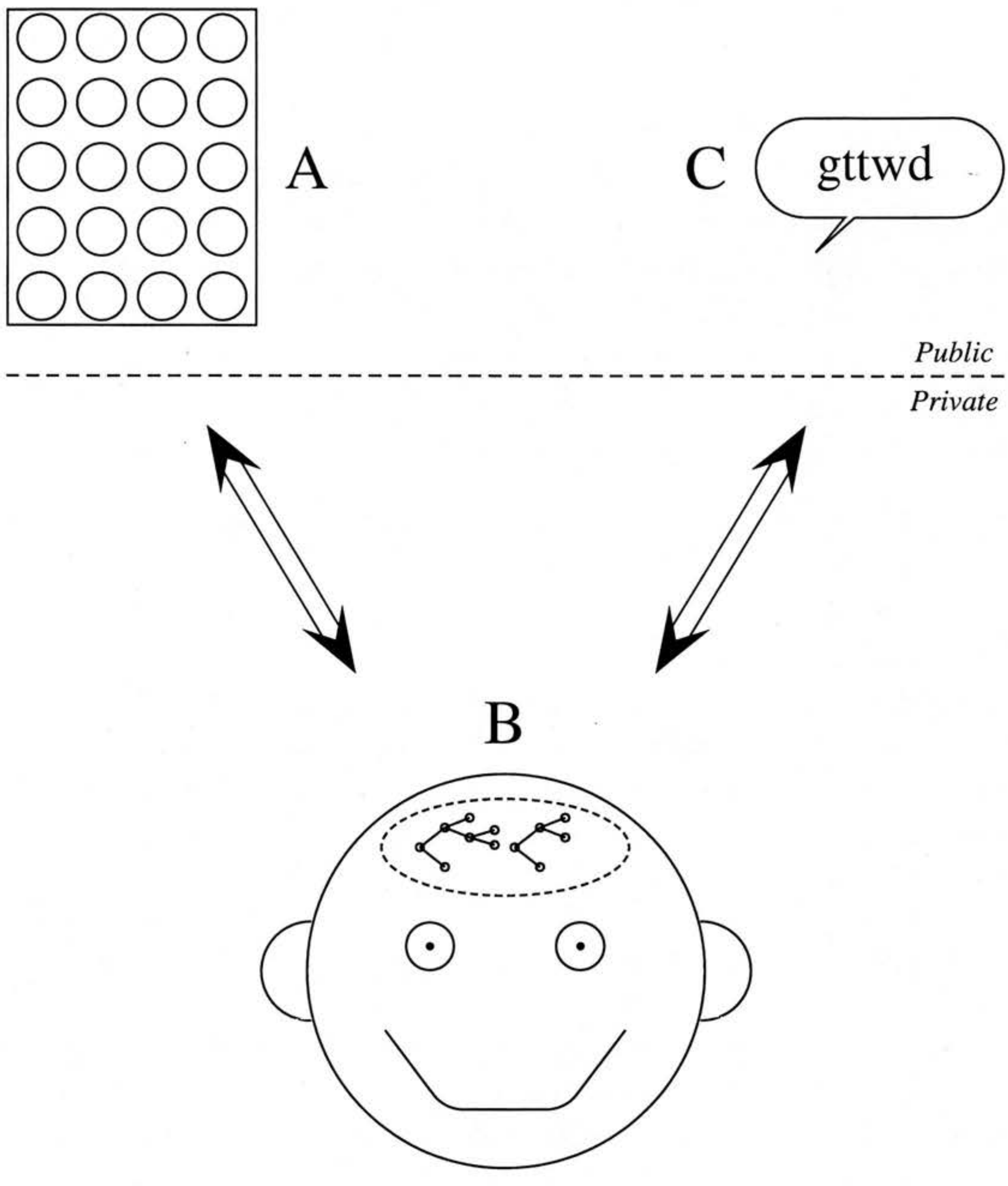


Figure 6.3: A model of communication which avoids explicit meaning transfer and the signal redundancy paradox must have three levels of representation for the agents: an external environment (A); an internal, private semantic representation, represented by the trees in the agent's brain (B); and public signals (C). The mappings between A and B, and between B and C, represented by the arrows, must also be private and inaccessible to other agents.

For instance, Hutchins and Hazlehurst (1995), in their famous neural network model of the development of shared vocabulary, present a model with an external world made up of events, or 'scenes'. At first glance, it is tempting to assume that because there is an external world, this model does not rely on explicit meaning transfer, and yet the scenes of the external world are not quite what they seem. The scenes are *themselves* used as the meanings for the agents; although the scenes are not explicitly transferred between the internal representations of the agents, they are still accessible to all the agents, and so they are still explicitly connected to particular signals during the communication process. In terms of figure 6.3, although there is an external world (A), there is no private semantic representation (B); the whole model takes place above the dotted line, and because there is no level of their model which is private to each agent, we return again to the signal redundancy paradox; if the agents know which scene is being talked about, there is no need to learn a signal.

Brighton (2002), too, presents a model which appears to contain an external world made up of communicatively relevant situations. Yet, as we saw in chapter 4, although the environment is defined by Brighton as the source of the meanings used by the agents, this relationship plays no role in the simulations; the agents never interact with the environment, and the mapping from communicatively relevant situations to meanings is pre-determined by the experimenter, and identical for all agents. Again, there is no private level in the model, and the environment, such as it is, is merely a complicating factor in the simulation, which is not able to solve the signal redundancy paradox.

At the beginning of this section, I presented Hauser (1996)'s summary of many different definitions of communication, from which he derived the crucial concepts of *information* and *signal*, in addition to the participants of the communicative episode. At that point, I tentatively assumed that the agents' meanings were the equivalent of Hauser's information, but it is now apparent that this view must be modified slightly. One possibility is to extend the notion of information so that it includes not only the internal semantic representations, but also the external experiences from which they are derived. It seems rather unsatisfactory, however, having established the importance of the three levels of representation in avoiding the signal redundancy paradox, to collapse private and public representations into the same, rather bland, category of information. It is more helpful, I feel, instead to extend the number of core communicative constituents, replacing information with meaning, and then additionally including the situations in the environment as another, fifth constituent of communication. This leaves us with a communicative episode which is made up of five essential elements:

- the participants of communication:
  - the speaker;
  - the hearer.
- and the components of communication:
  - a situation or object in the external environment, which serves as the subject of the communication;
  - a meaning, internal to the agent, which represents the subject of communication;
  - a signal, which is an external representation of the meaning.

In the above schema, the components of communication are explicity linked together through the referential nature of meaning (Frege, 1892), which is therefore crucially important to communication as a whole; the internal meaning must refer to a situation or object in the external world, and must itself be represented by the communicative signal. In the following sections, we shall look in detail at the process of communication without explicit meaning transfer, with particular emphasis on establishing a definition for successful communication, and the conditions under which this is most likely to occur, and thereby present one of the main contributions of this thesis[3].

6.3 Evaluating Communication

Having established the constituents of communication, I will now turn my attention to the evaluation of the communicative episode itself, and start by defining the following simple measure of communication which we will use throughout the experiments:

communicative success rate (κ) the percentage of successful communicative episodes.

Although this is a straightforward definition of the communicative success rate, a definition of communicative success itself is much more tricky, as we shall see. The evaluation of whether a communicative episode has succeeded or not can be rephrased in terms of whether or not the speaker and hearer have come to some agreement over the meaning of the signal; if the signal signifies the same thing to both speaker and hearer, then the

[3]Further work on the establishment of communication without using explicit meaning transfer can be seen in A. Smith (2001, 2003a).

communicative episode is a success. But how do we measure whether the speaker and hearer agree on this issue? If we focus in more detail at the details of communication, we realise that the speaker's role is to encode a particular meaning as a signal, and utter this to the hearer. An obvious starting point is to consider the episode a success, therefore, if the hearer, on decoding the signal into its internal meaning, arrives at the same meaning as that which was initially conceptualised by the speaker.

6.3.1 Sense Identity

Figure 6.4 shows a diagram of this kind of evaluation of the communicative episode, which I call an evaluation based on *sense identity*, because it compares the agent's internal meanings to each other. Two of the speaker's sensory channels are shown, with the discrimination trees which have been built thereon; the speaker's meaning is the shaded node in the lower half of the second tree. In the communicative episode, this meaning is encoded into the signal, and then decoded back by the hearer into its own meaning, denoted again by the shaded node in its meaning structure. The evaluation of the communicative episode is simply a matter of comparing the two nodes; if they are identical, then the agents are using the same meaning for the signal, and thus the episode is a success. Note that the overall semantic structures of hearer and speaker are not identical, because their trees have been grown in different places, but the particular node that has been used by both agents is in both structures.

Although a definition of successful communication based on sense identity is attractive and easy to conceptualise, there are problems with it which make it less than ideal for the evaluation task we require. Firstly, it requires that we look into the minds of the agents, to determine the meanings that they are using; although in the context of simulations, this is quite rightly possible and appropriate for the experimenter to do, it means that we are automatically ruling out the possibility that the agents themselves will ever be able to evaluate their own communicative episodes. Getting round this problem by allowing agents to read each other's minds, of course, throws us back into the signal redundancy paradox we are trying to avoid. Secondly, by focusing on sense identity as the evaluative mechanism, we are in effect saying that agents with slightly different semantic structures will never be able to communicate with each other accurately. For example, in figure 6.4, if the speaker had chosen a more specific meaning than the one shown, one level further down the tree, then this communicative episode could never have been a success, because the corresponding meaning structure does not yet exist in the hearer's meaning representation.

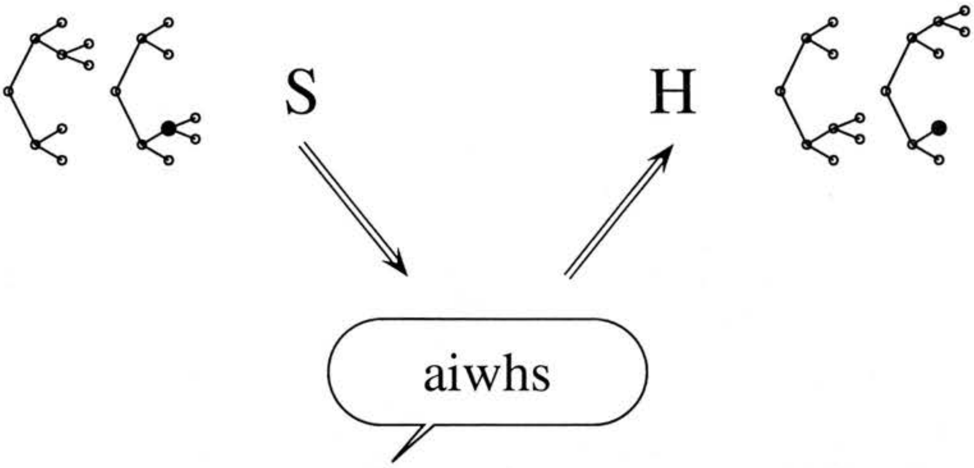


Figure 6.4: A model of communicative success based on sense identity. The speaker encodes a meaning S, denoted by the shaded node on the dendrogram, into the signal *ai-whs*. The hearer decodes the signal into its own meaning H. The communicative episode is evaluated by comparing S and H; if they are identical, as in this case, then it is a success, otherwise it is a failure.

Finally, it is worth noting that communicative success based on sense identity is vacuous if we allow meanings to be part of the linguistic transfer; it is inconceivable that the hearer would not produce the same meaning as the speaker, if that meaning was explicitly transferred to him! The final point should not come as too great a surprise, as we have already seen how explicit meaning transfer undermines the essential communicative nature of a model, but the previous arguments are more damaging, as they suggest unrealistic control over the exact specification of semantic structure or over access to other individual's internal workings, which undermines the important dividing line between public and private parts of the model. Evaluation of communicative success based on sense identity is possible, therefore, but not very desirable. How might we improve the situation, then, to avoid the pitfalls we have just described?

6.3.2 Reference Identity

The problems with evaluating communication in terms of sense identity can be summed up by the statement that such an evaluation is simply not realistic. If a person is trying to communicate with someone who speaks a foreign language, for instance, they *cannot* look into their head to see that their interlocutor has the same semantic structure as they do. If someone asks for a particular object using the word *jardal*, they can only gauge their success on whether or not their interlocutor passes them the object they had in mind,

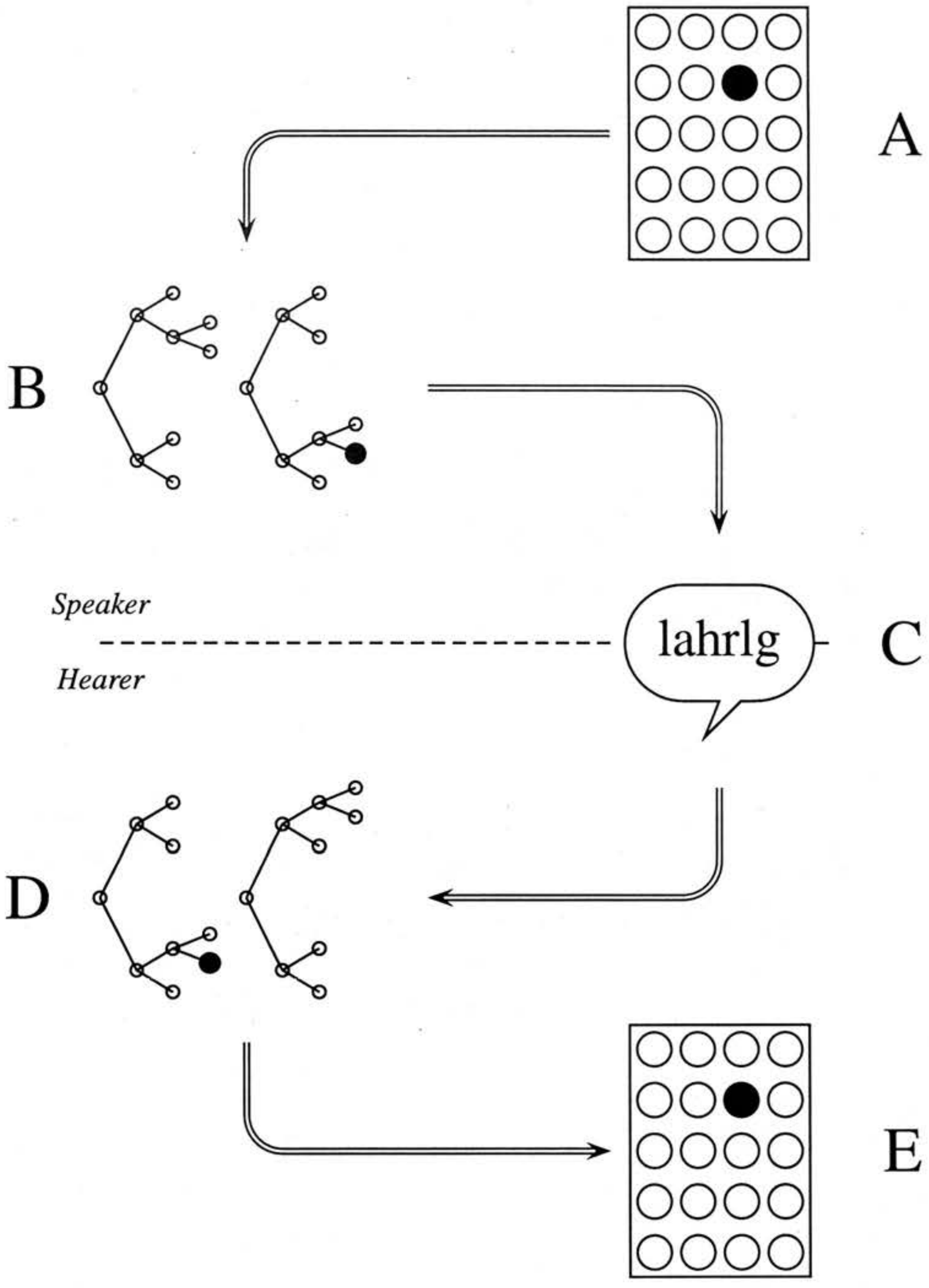


Figure 6.5: A model of communicative success based on reference identity. The speaker describes the shaded object at A with a meaning B, denoted by the shaded node on the dendrogram, and then encodes this meaning into the signal *lahrlg*, shown at C. The hearer decodes the signal C into its own meaning D, and then finds the object E which this meaning refers to.
The communicative episode is evaluated by comparing A and E; if they are identical, as in this case, then it is a success. Note that the communicative episode can succeed even though the meanings used by speaker and hearer (B and D respectively) are different; under sense identity this particular game would have been evaluated as a failure, since $B \neq D$.

not on whether she actually has the same semantic representation associated with the word *jardal* in her mind. In effect, we can only use what is available in the public domain (see figure 6.3) to evaluate the communicative episode indirectly, using the referents of the speaker's and hearer's meanings, rather than directly using their senses.

Figure 6.5 shows this situation diagrammatically: the speaker represents one object by a particular node on a discrimination tree, and then encodes this into a signal; the hearer then decodes the signal into a meaning, and then picks out the object to which that meaning refers. The evaluation of the communication episode using reference identity compares the objects, not the meanings themselves. In the particular example in figure 6.5, for instance, the agents use different meanings for the signal, but both these meanings still end up referring to the same object, so the communicative episode succeeds. Figure 6.5 also exemplifies one of the problems highlighted in the last section about sense identity evaluation: because the meaning node chosen by the speaker does not exist in the hearer's semantic representation, there is no way this communicative episode could succeed in terms of sense identity, even though both agents are using the same signal to refer to the same object.

6.3.3 Joint Attention

We have established that the most satisfactory way of evaluating a communicative episode is to use the referents of the speaker's and hearer's meanings, rather than by inspecting their internal meaning structure directly. This has a number of implications for the setup of the experiments, which it is useful to spell out explicitly. Firstly, and most obviously, the hearer must not know which object is being referred to; if the hearer is informed of the speaker's object, then there is no possibility of communication failure, and we return to a situation similar to that found in the signal redundancy paradox.

On the other hand, evidence from children acquiring language suggests that one of the most important building blocks used by children is *joint attention* with their interlocutors on the referent (Tomasello, 1999), where both child and adult attend to the same referent, and the child can infer, at some level, the intention of the adult to refer to the object. Baldwin (1991, 1993) shows in a number of related studies that small children are unable to learn words for objects simply by hearing the word while they are attending to the object; instead, learning can only take place if the child notices that an adult is explicitly directing their attention at the object, and that the adult is explicitly naming the object. Joint attention can occur through many different tasks and activities, though Tomasello provides a useful summary of these, and suggests that the three main types of interaction which result in joint attention are:

checking the attention of an adult in close proximity, by simply looking at them;

following the attention of an adult to a more distant object, perhaps by following their gaze;

directing the attention of an adult to a more distant object, by pointing at it.

The first of these interactions is clearly more fundamental and straightforward, as the child needs only to recognise that the adult is around and concerned with them; in following or directing an adult's attention, the child needs to know exactly what the adult is attending to. This appears to cause problems for our model of communication, in which, as we have seen, the hearer must not know what the speaker is attending to, lest the signal redundancy paradox reappear.

Despite the apparent importance of joint attention, there are some societies in the world where joint attention is explicitly excluded by cultural customs and traditions. In these cultures, adults speak to each other in the presence of children, but rarely name things for children, and direct speech towards children infrequently if at all (Lieven, 1994; Brown, 1998), until the children themselves have learnt enough of the language to be considered as interlocutors[4]. In such situations, it is clear that the children must decipher what the adults are talking about without the benefit of joint attention, but it is important to recognise, however, that children in these circumstances certainly learn words much more slowly than others who are not restricted in this way. Tomasello and Todd (1983), for instance, have shown that children who learnt with their parents using joint attention had larger vocabularies than those who did not, and word learning is also improved if adults name objects on which children have already focused their attention (Tomasello & Farrar, 1986).

We must, therefore, conclude that the establishment of joint attention is a very important mechanism which helps narrow down the possibilities in word learning, and allows words and language to be learnt more quickly, but on the other hand we should also prefer to develop a model in which joint attention is not a necessary condition for success, so that we can also account for the experiences of children who do not use joint attention or receive child-directed speech, and yet still acquire language successfully.

[4]In the case of the Tzeltal-speaking children studied by Brown (1998), this does not occur until the children are walking properly.

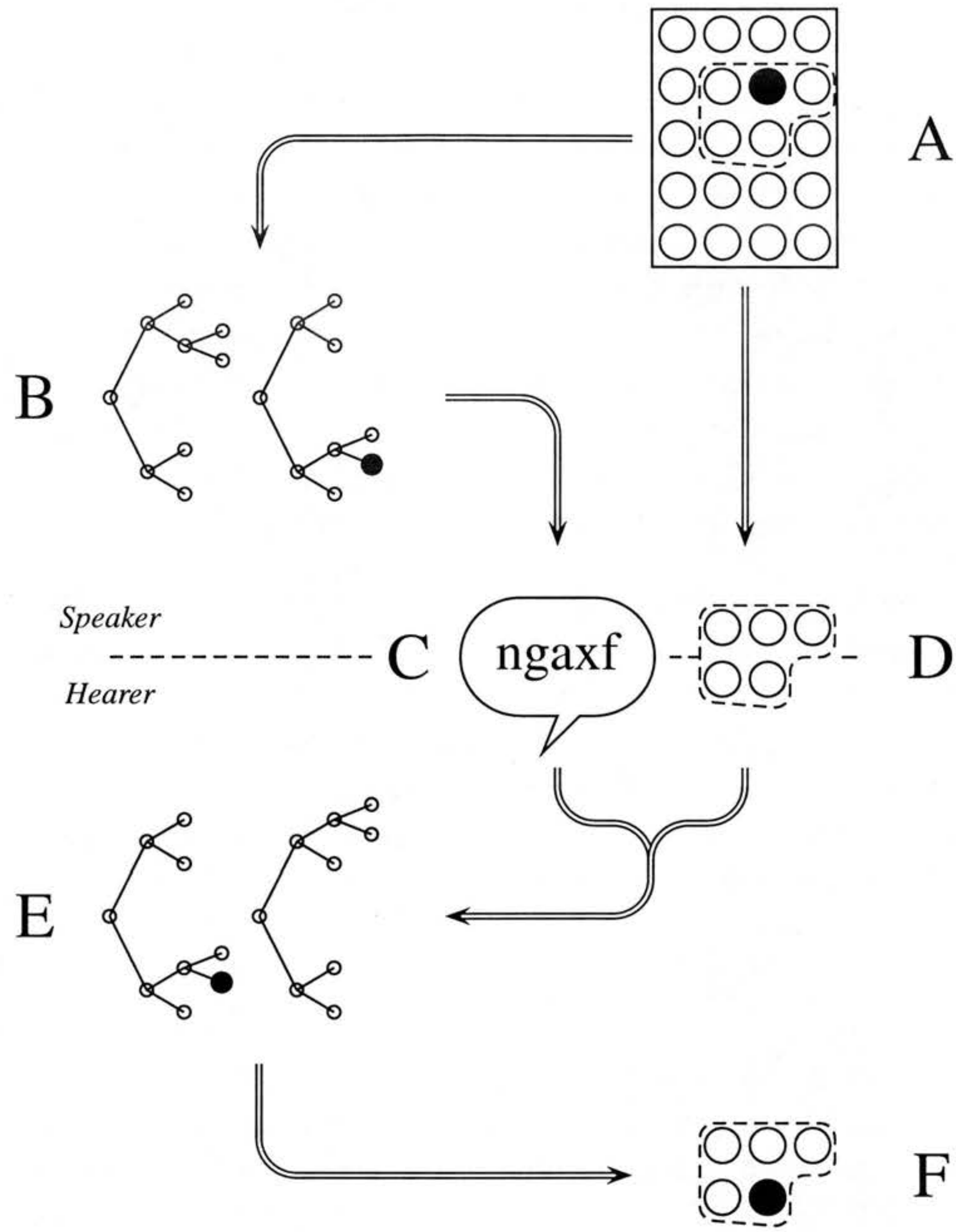


Figure 6.6: A model of communication based on reference identity and restricted joint attention. The speaker distinguishes the shaded object at A from the other objects in the context, shown within the dotted line, using a meaning B, and then encodes this meaning into the signal *ngaxf*, shown at C.
The hearer receives both C and the context in which the word was uttered, without knowledge of the original target object, shown at D; access to D establishes a restricted form of joint attention on the context between hearer and speaker.
The rest of the episode continues as in figure 6.5: the hearer decodes the signal into a meaning E, then finds an object F in the context to which E refers. The communicative episode is based on reference identity, and so is evaluated by comparing A and F. In the episode shown, the objects do not match, and communication fails.

Figure 6.6 shows the communicative model described in this thesis, which constitutes a position of restricted joint attention, in which the hearer and speaker both attend to the same *subset* of objects. The hearer is not being informed of the referent, and yet some form of joint attention is achieved on this subset, which eliminates other objects from the list of possible meanings. In order to obtain this subset of objects, we extend the role of the context from its original use in the discrimination game described in chapters 4 and 5. As can be seen in figure 6.6, the speaker is given a random context of objects, from which one is chosen as the target object, and it finds a distinctive category which both identifies the target object and does not identify the other objects in the context, which it represents by the utterance *"ngaxf"*. In the communicative episode, the hearer receives not only the utterance, but also the context in which the word was chosen. Both hearer and speaker are therefore jointly attending to the same group of objects.

It is important to note, however, that the hearer's context (shown as D in figure 6.6) differs from the speaker's context (shown inside the dotted line within A in figure 6.6), in that the identity of the target object is kept secret from the hearer. This is necessary because of our use of reference identity to evaluate communicative success. Absolute joint attention, where there is just one object in the context, eliminates all the uncertainty in the hearer's mind, and will inevitably result in perfect 'communication', albeit communication which once again suffers from the signal redundancy paradox.

6.3.4 Reinforcement and Feedback

Another strategy which works very well in associative learning is reinforcement learning, which has been widely used in simulations of the acquisition of vocabulary, for instance in the *guessing games* described, for example, by Steels and Kaplan (2002) and by Vogt (2002). Under reinforcement learning, agents receive an evaluation of their actions, which they use to modify their future behaviour. For example, if they use a word 'correctly', they will receive some kind of reward, or positive feedback, to encourage them, but if they use a word 'incorrectly', they will receive a disincentive, or negative feedback, and will also usually be provided with the correct word they should have used.

In chapter 3, however, we saw that although both positive and negative feedback do appear to be used by parents in 'Western' cultures, during child-rearing, this is not by any means a cultural universal. We have already looked at Lieven (1994)'s description of cultures in which parents do not even speak to their children in the initial stages of acquisition, much less provide them with feedback, and at Bloom (2000)'s description of the study on mute children who developed normal language in the absence of the possibility of feedback.

The receipt of feedback by the learner, therefore, should be treated with severe caution and, I would contend, should not be assumed by a model of the acquisition of communication systems. One important result from my thesis, indeed, will show that successful communication can be modelled without the need to provide feedback of either kind to the agents.

6.4 Communicative Strategies

In this section, I will firstly describe previous work looking at the evolution of communicative strategies in simulated populations, and will then go on to present in some detail the communicative model on which the simulations described herein are based. The earliest experiments into the evolution of communicative strategies were carried out by Hurford (1989); from a communicative point of view, Hurford's model suffers from similar problems as did Hutchins and Hazlehurst (1995)'s, in that the meanings are objects in the external world which are accessible to all agents and there is no private domain, but his model still provides important insights into the nature of successful communicative strategies.

Hurford introduces the notion of two dynamic communicative matrices for the agents, with one type of matrix to encode their *transmission* behaviour, or the probability of uttering a signal for a particular meaning, and the other to encode their *reception* behaviour, or the probability of interpreting a meaning for a particular signal. Each agent has a fixed life-cycle of birth-acquisition-reproduction-death, and the focus of the experiments is the acquisition stage, in particular the acquisition of different communicative strategies which the agents use to form their two communicative matrices. In each case, the agent is given small samples of the communicative behaviour of the previous generation, averaged across the whole population of adults into population matrices: for each meaning, one signal is chosen probabilistically from the population transmission matrix; and for each signal, one meaning is chosen probabilistically from the population reception matrix.

There are two different ways used by the agents to form their own new matrices from these sample matrices: direct copying and indirect optimisation. If the matrix being formed is of the same type (transmission or reception) of matrix as the source, then the new matrix is copied directly from the sample; if an agent observes signal s_1 being interpreted as m_1 in the sample reception matrix, then the agent will set the probability of interpreting s_1 as m_1 to 1.0 in its own reception matrix. If, on the other hand, the new matrix is of a different type, then it is not copied, but instead optimised indirectly to reflect the sample. For example, if two signals s_1 and s_2 are both interpreted by meaning m_1 in the sample reception matrix, then a derived *production* matrix would produce both signals with an equal probability of 0.5 for meaning m_1. Hurford defines three different learning strategies, according to which of these formation rules the agent uses for each of its matrices:

Imitators copy both matrices.

They form their transmission matrix from the sample of transmission behaviour, and their reception matrix from the sample of reception behaviour;

Calculators optimise both matrices.

They form their transmission matrix from the sample of reception behaviour, and their reception matrix from the sample of transmission behaviour;

Saussureans copy their transmission matrix, and optimise their reception matrix.

They form their transmission matrix from the sample of transmission behaviour, and their reception matrix from *their own* transmission matrix.

Hurford (1989) shows that the Saussurean agents are the most successful at developing communicative systems, while Calculators are unable even to maintain a communication system with which they are provided. The Saussureans' success comes from their optimisation of one behaviour from the other, and in particular ensures that the agents have a *bidirectional* mapping between signal and meaning such as that presented as the most fundamental linguistic structure by Saussure (1916). Hurford shows that the bidirectionality of the communicative mapping between signals and meanings has clear advantages over the other systems. The Calculators, on the other hand, try to optimise both their behaviours at once, and crucially do not attempt to synchronise their own behaviour into any form of bidirectionality, and their system fragments and drifts into chaos as a result. The most important insight of this work is that, by separating transmission behaviour from reception behaviour, Hurford has shown that *lexical bidirectionality*, or the coupling of these two behaviours is crucially important for the development of successful communication systems.

A similar model, which is also based on agents using communicative matrices to control their behaviour, is presented by Oliphant and Batali (1997). Both the agents in this and in Hurford (1989)'s model try to accommodate the transmission and reception behaviour of the whole population in determining their own communicative matrices, but there is one important difference. While Hurford's agents use a probabilistic sample of the population matrix, Oliphant and Batali's agents use the average population matrices in their entirety to determine their communicative strategies. In the latter model, the agents choose a winner-take-all algorithm, choosing the behaviour which is most popular for each signal or meaning respectively, which necessarily results in only one signal being chosen for each meaning, and only one meaning being chosen for each signal. Within this framework, Oliphant and Batali define two different learning strategies, as follows:

Imitate-Choose The agents form their transmission matrix as follows:

- for each meaning, find the signal which is sent most often in the population.

and their reception matrix as follows:

- for each signal, find the meaning which is interpreted most often in the population.

Obverter The agents form their transmission matrix as follows:

- for each meaning, find the signal which is interpreted as the meaning most often in the population.

and their reception matrix as follows:

- for each signal, find the meaning which is sent with the signal most often in the population.

In this way, the Imitate-Choose algorithm bases its behaviours *directly* on the relevant population matrices, while the Obverter algorithm bases its behaviours on them *indirectly*; its transmission matrix is based on the population's reception behaviour, and its reception matrix is based on the population's transmission behaviour. Essentially, a speaker using Imitate-Choose chooses the word it knows other agents use to represent the target meaning, while a speaker using Obverter chooses the word it knows other agents will understand as the target meaning.

Oliphant and Batali show that the Imitate-Choose algorithm can increase communicative accuracy, but only in a model with a system which already has a high level of communicative accuracy; in effect, it polarises the system further, improving when the system is good and degrading when the system is bad. Obverter, on the other hand, always produces steady increases in communicative accuracy until an optimal system is reached. Oliphant and Batali suggest that this happens because of the implicit avoiding of ambiguity in the obverter mechanism, and the fact that an attempt to communicate is built into the very choices the agents make, because the agents base their production matrix on the population's reception matrix. K. Smith (2001, 2002b), shows, however, in neural network models of communicative agents, that some agents who base their production matrix on the population's *production* behaviour can also still develop an optimal communication system. After further analysis of his system, K. Smith argues that, rather than the design of the obverter mechanism itself being responsible for the establishment of optimal systems, there is actually a key bias which favours one-to-one biases between meanings and signals. He goes on to show convincingly not only that Hurford (1989)'s original model, and Oliphant and Batali (1997)'s subsequent model contain this bias, but also that many other models proposed in the literature to explain the cultural evolution of communication, including those presented by Batali (1998), Kvasnička and Pospíchal (1999), Livingstone and Fyfe (1999) and Kirby and Hurford (2002), also contain this very same key one-to-one bias.

Crucially, K. Smith (2002b) suggests that there are two pre-conditions for the emergence of optimal communication: the key bias in favour of one-to-one mappings between meanings and signals; and the capacity to read other people's intentions. These two conditions are strongly related to phenomena we have explored already in this thesis. The key bias is, of course, strongly supported by the proposed existence of similar biases which attempt to explain the acquisition of vocabulary by children, such as Clark (1987)'s Principle of Contrast and Markman (1989)'s Mutual Exclusivity Bias, which we explored in chapter 3. His second condition, however, the ability to read other's thoughts, is more problematic, and yet also crucial to the emergence of communication, because much depends on the nature of this proposed capability. On the one hand, if intentions can be completely perceived, then we are led back to the iniquitous signal redundancy paradox we explored and rejected earlier in this chapter. On the other hand, we have already acknowledged the importance and usefulness of joint attention in providing a restricted context to solve the problem of the indeterminacy of meaning, while noting that some human cultures appear to manage without it. If the capability of reading others' intentions can be restricted to providing the context alone, then we can surely reconcile the difficulty.

In this section, we have explored many simulations which have shed light on what is needed to develop a stable communication system between agents. Hurford (1989) has shown the importance of lexical bidirectionality, the fact that reception and transmission behaviours are coupled together. (Oliphant & Batali, 1997) have proposed an obverter learning algorithm, which always constructs an optimal communication system, and K. Smith (2002b) has proposed that obverter and other successful algorithms are underpinned by a key bias in favour of one-to-one mappings between signals and meanings. In the rest of this chapter I shall explain the workings of the lexicon in my model, which incorporates both lexical bidirectionality and aspects of an obverter system of learning. This model will not require the agents to be able to read each other's minds, but will include a truly external world in which to ground the meanings and which is necessary to avoid the system redundancy paradox which exists in all the computational models which have been hitherto suggested.

6.5 The Lexicon

The communicative model as described hitherto contains at least two participants, and at least three levels of representation split across two different domains, that which is public and accessible to all, and that which is private and individual to each agent. The communicative process in the model is made up of three distinct parts:

signal production: the speaker, having found a distinctive category in the discrimination game, chooses a signal to represent this meaning.

signal transfer: the signal is transferred to the hearer in conjunction with the whole context in which the signal was chosen.

It is important to remember, with respect to signal transfer, that the speaker's meaning is *not* transferred, nor does the hearer know which object in the context is being referred to.

interpretation: the hearer tries to interpret the signal in the context in which it is heard.

Crucial to both the production of signals by the speaker and their later interpretation by the hearer, and therefore at the heart of the communication process, is each agent's individual set of private mappings between the meanings stored on its discrimination trees, and the utterances used to describe them. By implementing both production and interpretation from the same set of linguistic mappings, I am ensuring that lexical bidirectionality, of the kind explored and shown to be crucial by Hurford (1989), is necessarily present in the system. In this section, I shall be focusing in detail on the mapping, which is stored as a dynamic lexicon of associations between words and meanings; each entry in the lexicon contains the following components:

- a signal s;
- a meaning m;
- a count of how many times the pair has been used u;
- a confidence probability p, which represents the agent's confidence in the association between the signal and meaning.

The properties of both signal and meaning are clearly straightforward, but a few explanatory words are necessary with reference to the usage count and the confidence probability. A signal-meaning pair can be used both by being uttered by the speaker and by being understood by the hearer, so that u is the total number of communicative episodes in which the agent either uttered s to represent m, or interpreted s as representing m. An agent's confidence in a signal-meaning pair is based solely on the relative co-occurrence of signals and meanings, or the proportion of times in which s has been used that it has been associated with m. More formally, $p(s, m)$ can be expressed as:

$$p(s,m) = \frac{u(s,m)}{\sum_{i=1}^{l} u(s,i)} \tag{6.1}$$

where l is the number of entries in the lexicon. This confidence probability represents the agent's interaction history, recording, in summary form, all the associations between signals and meaning which the agent has ever made, and the equation above is equivalent, as has been pointed out by Vogt and Coumans (2003), to the conditional probability that, given a particular signal s, the meaning m can be expected.

A short extract from an example lexicon is given in table 6.1, and this extract will be used in the following sections to explain how the algorithms for choosing words and meanings work. Each agent's complete lexicon is obviously potentially very big, with a potential size of $S \times M$ entries, where S is the total number of signals, and M the total number of meanings for the agent, and grows considerably over the length of an experiment, depending on how frequently the agents create new words, so table 6.1 shows only the entries for two particular signals (*gttr* and *oij*), and the meanings associated with them. For reference, however, I have provided examples of an agent's complete lexicon in appendix E.

Signal	Meaning	Usage	Conf. Prob.
gttr	0-0	1	0.083
gttr	0-1	2	0.167
gttr	0-11	1	0.083
oij	1-0	9	0.600
gttr	2-0	4	0.333
oij	2-0	6	0.400
gttr	2-1	1	0.083
gttr	3-1	2	0.167
gttr	4-00	1	0.083

Table 6.1: An extract from an example lexicon. Each entry contains a signal s, a meaning m, a count of usage u and a confidence probability p.

6.5.1 Signal Choice

Given the lexicon extract in table 6.1, how does the speaker decide which signal to choose, when it is trying to express a particular meaning? Let us assume, for argument's sake, that the speaker has played a discrimination game, and found a distinctive category $2-0$ which distinguishes the target object from the other objects in the context. There are a number of different algorithms which the speaker could implement in order to choose a word to represent $2-0$, as we have seen in the previous section. Remember, however, that in the previous models the agents had access to the production and reception matrices of all other agents, either through access to the combined population matrices of Oliphant and Batali (1997) or the sampled population matrices of Hurford (1989). Having rejected this, the only lexicon to which the agent has access is actually his own, so I assume that the lexicon shown in table 6.1 is the only lexicon on which the agents will base their transmission and reception decisions.

Returning to the task in hand, the signal *oij* would seem to be a reasonable choice to represent $2-0$, based on the lexicon shown in table 6.1 for two obvious reasons:

- the value of u is higher for the signal-meaning pair [*oij, 2-0*], namely 6, than for [*gttr, 2-0*] (4); *oij* has therefore been used in association with $2-0$ more often than *gttr*.
- the value of p is higher for the signal-meaning pair [*oij, 2-0*], namely 0.4, than for [*gttr, 2-0*] (0.33); the agent is therefore more confident in the association of *oij* with $2-0$ than it is in the association of *gttr* with $2-0$.

Signal s	Usage $u(s,m)$	Conf. Prob. $p(s,m)$	Interpretation $i(s)$
gttr	4	0.333	2-0
oij	6	0.400	1-0

Table 6.2: The introspective obverter strategy. Agents use the lexicon extract shown in table 6.1 to choose a word to represent the meaning m $(2-0)$, by finding a list of candidate words (here *gttr* and *oij*), which are shown together with their usage in association with m, $u(s,m)$, the confidence the agent has in their association with m, $p(s,m)$, and the meaning which they would be interpreted by the agent, $i(s)$, or the meaning which maximises the confidence probability for the signal s. The agent would interpret *gttr* with the appropriate meaning, but not *oij*, despite the fact that both u and p are higher for the latter word.

Choosing *oij* on this basis would in fact be similar to using the Imitate-Choose algorithm described by Oliphant and Batali (1997), in which the agents searches through the signals in the population transmission matrix, choosing the one which is most popular, if we assume that the lexicon is table 6.1 represented the population lexicon. However, Oliphant and Batali's own results have already demonstrated that the Imitate-Choose algorithm polarises a communication system it is imposed on, improving only systems which are already high in communicative accuracy. Obverter, on the other hand, gradually increases the communicative accuracy of a population over time until an optimal system results. I will, therefore, base this model on a modified version of the obverter strategy, but one which avoids mind-reading and explicit meaning transfer.

Unfortunately, true obverter learning as described by Oliphant and Batali assumes that the speaker can read the lexicons of the other members of the population, to calculate the optimal signal to use for any meaning, and thus allow the speaker to choose words which he knows the hearer will understand correctly. We have seen already how such mind-reading is not only unrealistic, but more damagingly returns us to the telepathic world of the signal redundancy paradox, and so I assume instead that the speaker has access only to its own lexicon, using this alone as an approximation to the general population lexicon and as a basis for decision-making. Instead of explicitly choosing the word that will be understood most generally in the population, the speaker using the *introspective obverter* strategy chooses the word that *it itself* would be most likely to understand if it was the hearer.

In order to decide which word the agent should choose using introspective obverter, we need to investigate how the two candidate words *oij* and *gttr* would be interpreted. The precise details of signal interpretation will be discussed in more detail below, but for expository purposes here, it will be sufficient to say that the confidence probability is the crucial statistic on which a hearer decides what a word must mean; given the uncertainty of meaning inherent in the system, it chooses the association in which it has the most confidence. In order to see how a word is interpreted, we need to find the meaning which maximises the confidence probability for each word. We can see from table 6.1 that although *oij* has been associated with the meaning $2-0$ on more occasions than *gttr*, it would actually be interpreted as $1-0$, because $1-0$ is the meaning which maximises the confidence probability for *oij*, while *gttr* would be interpreted with the target $2-0$ meaning, as this is the meaning which maximises the confidence probability for *gttr*. Table 6.2 gives a summary of how the introspective obverter strategy allows the speaker to choose a word that the speaker would understand if it was the hearer.

Interestingly, the agent would not find a word from its lexicon in table 6.1 to express many meanings which do nevertheless have some associations (e.g. $0-0$, $3-1$ etc.). One of the characteristic outcomes of obverter learning is the avoidance of homonymy, so we find that, at any one time, each word in the lexicon is only used with one meaning, although the particular meaning can of course change as the associations in the lexicon are updated. This means that, although there are eight meanings in the lexicon extract, only two of them are actually used by the speaker, and so only these can be regarded as being truly in the speaker's *active lexicon*.

6.5.2 Meaning Choice

We have seen how the speaker tries to second-guess the hearer and chooses words which are likely to be understood before uttering them, but a much greater problem is faced by the hearer in understanding the meaning which is being conveyed. On hearing a signal, the hearer's only guide in determining the intended meaning, in addition to the signal itself, is the observation of the context in which the word was heard. In figure 6.6, we saw that the hearer knows neither the target object to which the speaker is referring, nor the meaning which the speaker has in mind for the signal, although the restricted joint attention we have implemented means that there is only a subset of possible objects in the context to which the signal could refer. Despite this, the hearer tries to infer the intended meaning solely from the context and from its own previous experiential history, stored in its lexicon as described above, disambiguating the potential referents in the context as follows:

The hearer creates a list of *possible meanings* or *semantic hypotheses*, namely every meaning in its conceptual structure which identifies any one of the objects in the context and distinguishes it from all other objects in the context. All these possible meanings are equally plausible, and the hearer has no immediate reason to prefer one over the others, so each of them is paired in turn with the signal and lexicalised. The lexicalisation of a signal-meaning pair is carried out by incrementing $u(s, m)$, and recalculating $p(s, m)$ based on the new value of $u(s, m)$. Once all the possible meanings have been lexicalised, the hearer searches through its semantic hypotheses, and chooses the meaning in which it has the highest confidence; if it has an equally high confidence in more than one meaning, then one of these is chosen at random. The hearer then returns to the context, to find the object therein which is identified by the meaning it has just inferred. Because evaluation of communicative episodes is based on reference identity, this object is compared with the original target object of the speaker, to determine the success of the game, as shown in figure 6.6; successful communication occurs when the speaker's original target object is the same object as that which is identified by the hearer's meaning. It is not necessary that the agents use the same agent-internal meaning, only that both agents *refer* to the same object.

Neither agent, however, receives any explicit information about the success or failure of the episode. It is possible, however, especially if the hearer has very little conceptual structure, that its set of semantic hypotheses is empty. This means that there is no meaning in the hearer's conceptual structure which distinguishes any one of the objects in the context from all the others. This is analogous to being shown five identical mugs, and being asked to find the "fipply" one; it is impossible to interpret the signal "fipply", because there are no distinguishing features on which to make a decision, and the hearer cannot interpret the signal. In chapter 9, I explore what happens when meaning creation is driven not just by playing discrimination games, but also by failure to interpret unfamiliar signals.

6.6 Summary

In this chapter, I have described in detail the constituents which must make up an accurate model of communication, and have shown that one of the most important factors is that there must be a distinction drawn between public knowledge of events and objects in the environment on the one hand, and private semantic representations which only the individual agent can access on the other. I then discussed various potential ways of evaluating the success of a communicative episode, and showed that a system of reference-based identity is much more plausible, and potential evaluable by the agents themselves, than a system of sense-based identity. We saw from the work of Hurford (1989) how lexical bidirectionality is important in the evolution of vocabulary, and we learnt from K. Smith (2002b) that optimal communication systems arise if there is an underlying bias in favour of one-to-one biases between signals and meanings. Such a system, in which the speaker's choice of signal is informed by his knowledge of the hearer's lexicon, is described by Oliphant and Batali (1997) . In order to avoid enabling telepathic communication between the agents, I described a modification to this system, called *introspective obverter*, which allows agents to communicate without explicit meaning transfer, without knowledge of the topic of conversation, and without feedback about the communicative process itself. Examples of the lexicons which are built by the agents in the model are shown in appendix E for reference.

In the following chapters, I will describe experiments which investigate the conditions under which communicative success is likely to occur, showing that there is a close relationship between communicative success and meaning similarity, and I will then go on to describe experiments which investigate how the agents build individual conceptual structures which are more co-ordinated with those of their interlocutors.

CHAPTER 7

Preliminary Communication Experiments

> "Make your contribution as informative as required. Do not make your contribution more informative than required." (Grice, 1975, p. 45)

7.1 Introduction

The communicative model I described in chapter 6, using the introspective obverter communicative strategy on top of a system of grounded meaning creation, fulfills the initial objective of constructing a system of communication which relies neither on the explicit transfer of meaning, nor on feedback to guide the learning. Creating a framework for communication is clearly only the first step towards a simulated model of communication, however; the crucial question to ask about this theoretical model of communication is the degree to which it actually works.

- Can the agents within such a system communicate with each other with an acceptable level of accuracy?

In this chapter, I will thoroughly investigate the model's communicative efficacy; in section 7.2, I describe how meaning construction can be decoupled from communication, by providing the agents with pre-defined conceptual structures of various sorts, and then in section 7.3, I explore how successful agents are in developing a communication system under these circumstances, and look at how the level of meaning similarity between agents relates to level of communicative success they achieve. Section 7.4 explores why the level of meaning similarity in randomly provided conceptual structures is so predictable, while section 7.5 explains why communicative success is often higher than meaning similarity.

7.2 Innate Concept Provision

In this section, I will temporarily sever the link between meaning creation and communication, so that we can focus on communication alone, and see whether the framework we have built is viable. To do this, we must therefore temporarily dispense with the meaning creation algorithms, and instead give the agents innate conceptual systems[1]. There are two ways in which I provide the agents with innate conceptual structure, as detailed below:

comprehensive innate concept provision, where a discrimination tree is comprehensively refined to a certain depth, known as the *comprehensive depth*;

random innate concept provision, where a leaf node λ is randomly chosen on a discrimination tree and refined, or split into two sub-categories.

Figure 7.1 demonstrates the nature of comprehensive innate concept provision on a discrimination tree. Assume first that the root of the tree is at depth 0, and every movement down the tree away from the root increments the depth level. A tree is comprehensively refined when the following condition has been met:

- if one node of depth d has been refined, then every node on the tree of depth d must also have been refined.

If all nodes at depth d have been refined, then we say that the tree has a comprehensive depth of $d+1$. The tree on the left of figure 7.1, for example, has a comprehensive depth of 1, because the root node (at depth 0) has been refined; the tree in the middle has been comprehensively refined to a depth of 2, because both nodes at a depth level of 1 have been refined; likewise, the tree on the right has been comprehensively refined to a depth of 3, because each node at the previous depth level has been refined. Comprehensive tree growth covers the whole of the feature space (hence the name) at each level, and no part of the feature space has any more distinctions than any other. Note that there is only one way to comprehensively refine a discrimination tree to a certain depth, so if two trees have comprehensive concept provision of the same depth, the similarity τ (see equation 5.13) between them will always be 1.0. This, therefore, provides us with an easy way to guarantee that two agents will have synchronised meaning structures ($\sigma = 1$); we simply

[1]No claim at reflecting reality is involved here; it is merely experimentally convenient to be able to set up simulations where the meaning similarity σ between the agents can be explicitly controlled.

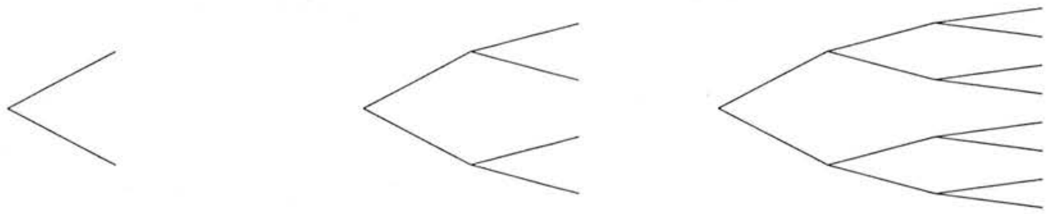


Figure 7.1: Comprehensive innate concept provision occurs when every node at a certain depth of the discrimination tree has been refined. On the leftmost tree, the node at depth 0 (the root) has been refined, so the tree has been comprehensively refined to depth 1. On the tree in the middle, both nodes at depth 1 have been refined, so the tree has been comprehensively refined to depth 2. On the rightmost tree, all nodes at depth 2 have been refined, so the tree has been comprehensively refined to depth 3.

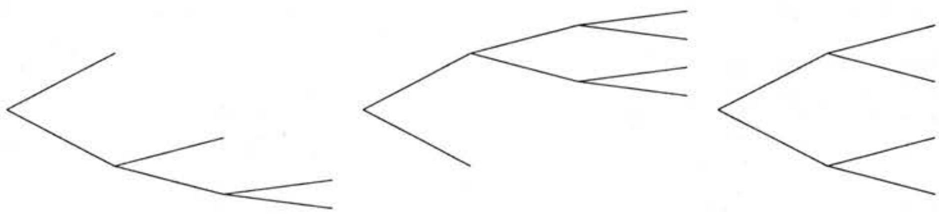


Figure 7.2: Random innate concept provision involves the random refinement of nodes on the agent's discrimination trees. The agent above has three sensory channels, which have been refined in total ten times, giving the agent an innate endowment of twenty different meanings, each represented by a node on a tree.

ensure that they are both provided with comprehensive conceptual structure to the same depth on each of their sensory channels. Likewise, if we vary the number of channels on which comprehensive growth is provided, then we can easily specify a particular level of meaning similarity, as we shall see in section 7.3.2.

Random innate concept provision, on the other hand, is not explicitly deterministic, but instead allows us to introduce an element of uncertainty into the proceedings; it is characterised by the simple choosing of a node on a tree at random, and then the refinement of that node. This selection and refinement procedure can of course be carried out repeatedly; figure 7.2 shows an agent with three sensory channels, which have been refined respectively three, four and three times.

7.3 Innate Concepts and Communicative Success

Using these methods of innate concept provision gives us a straightforward framework in which the level of meaning structure σ between agents can be pre-specified, allowing us to explore the effects that different kinds of conceptual structure have on the communicative success rate κ that the agents achieve. In this section, I investigate whether agents with innate meaning structure can communicate with each other in the model as it stands. A simulated world is created, in which there are twenty randomly generated objects, which are described in terms of five features. There are two agents, each having five corresponding sensory channels on which its discrimination trees are constructed; the actual details of the meanings we will provide them with, and the effects these have on communication, are the focus of these experiments. In these initial experiments, one agent acts as the speaker in the communicative episodes, and the other as the hearer; in effect, therefore, we are investigating whether the hearer can learn a mapping between signals and meanings which has been created by the speaker, without being told the meanings, without being told the referent of the communicative episode, and without any feedback on the success or otherwise of the learning process.

7.3.1 Synchronised Comprehensive Conceptual Structure

Firstly, we look at two agents who have been provided with innate conceptual structures with a comprehensive depth of 3 (see figure 7.1) on all their five sensory channels; because all their sensory channels have exactly the same level of comprehensive innate conceptual structure to the same depth, their meaning structures are necessarily synchronised ($\sigma = 1$). Figure 7.3 clearly shows that very high levels of communicative success κ occur under these conditions; in all cases, there is an initial sharp rise in κ, as the hearer deduces the meanings of many signals through their disambiguation in different contexts. The value of κ is already very high after only a few hundred communicative episodes, and after this initial rise, κ continues to climb more slowly, as the hearer tries to deduce the meanings of the remaining signals; this occurs because the meanings which these remaining signals represent are seldom needed in the discrimination games, and so occur relatively infrequently in communicative episodes, making them more difficult to disambiguate through exposure in different contexts.

In order to quantify the communicative success the agents achieve under these experimental conditions, each simulation is run 50 times, after which I calculate the average communicative success rate achieved after 5000 communicative episodes ($\bar{\kappa}$) and the coefficient of variation ($\text{CoV}(\kappa)$), which is the standard deviation expressed as a percentage

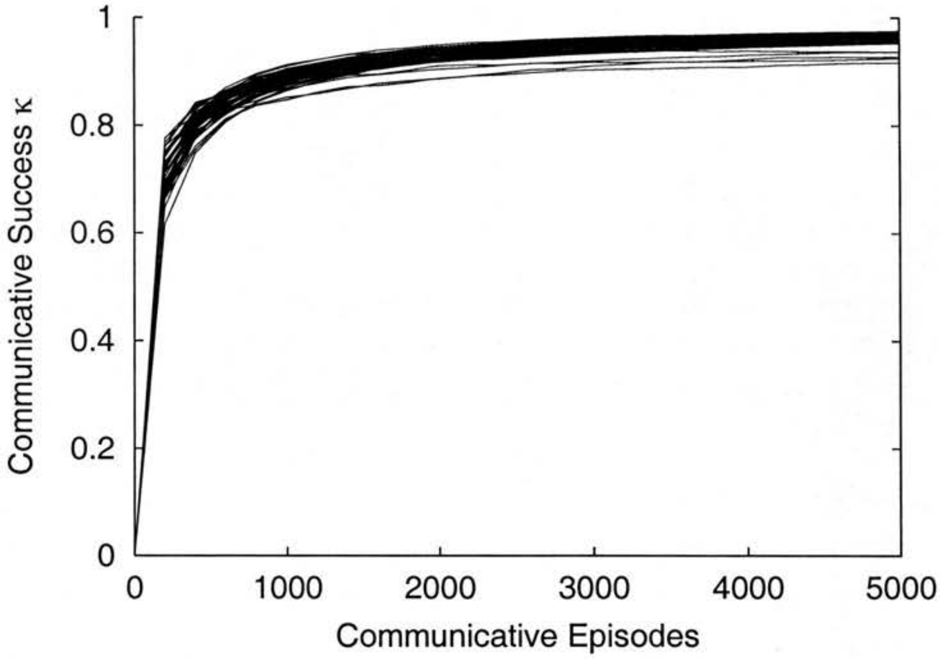


Figure 7.3: Communicative success κ in agents with initially synchronised ($\sigma = 1$), comprehensive, innate conceptual structures. Agents are each provided with innate conceptual structure to a comprehensive depth of 3 on each of their sensory channels. The experiment was repeated 50 times, with each line on the graph representing a single run.

Communicative Success Rate				
Mean		Range		
$\bar{\kappa}$	(CI)	Max(κ)	Min(κ)	CoV(κ)
0.96	(0.96 – 0.96)	0.98	0.92	0.01

Table 7.1: Communicative success κ in agents with initially synchronised ($\sigma = 1$) innate conceptual structures, provided to a comprehensive depth of 3. The table provides a summary of the range and distribution of κ after the 50 runs of the experiment shown in figure 7.3, showing the mean communicative success rate $\bar{\kappa}$, together with a 95% confidence interval around the mean, the range of values for κ, and the coefficient of variation.

Comprehensive Depth	Meaning Similarity	Communicative Success Rate				
		$\bar{\kappa}$	(CI)	Max(κ)	Min(κ)	CoV(κ)
1	1.00	1.00	(1.00 – 1.00)	1.00	1.00	0.00
2	1.00	0.99	(0.99 – 0.99)	1.00	0.98	0.00
3	1.00	0.96	(0.96 – 0.96)	0.98	0.92	0.01
4	1.00	0.95	(0.95 – 0.96)	0.97	0.91	0.01

Table 7.2: Meaning similarity σ and communicative success κ in agents with innate synchronised conceptual structure. Agents are innately provided with identical meaning structures ($\sigma = 1$) to various comprehensive depths, and then undertake 5000 communicative episodes. Each simulation is repeated 50 times, over which the mean communicative success rate $\bar{\kappa}$, together with a 95% confidence interval around the mean, the range of values for κ, and the coefficient of variation, are calculated.

of the mean, and gives a measure of the relative dispersion or variation at the cut-off point of 5000 episodes[2]. I express the mean together with a 95% confidence interval, because the particular 50 runs of the simulations whose results are shown here only represent a sample of all the possible runs which could have occurred. For completeness in viewing the range of the results, I also include the maximum and minimum values achieved for κ during the experiments. Table 7.1 shows these values for the experiment shown in figure 7.3, and confirms the results we can obtain by visual inspection of the graph, namely that the rate of communicative success after 5000 communicative episodes is very high ($\bar{\kappa} = 0.96$), and there is very little variation in the levels of communicative success (CoV($\kappa = 0.01$)) which are achieved.

The experiments were then modified slightly, so that the comprehensive depth of the innate conceptual structure on each of the agents' sensory channels was varied, and each experiment was again repeated 50 times, and summary statistics calculated. We can see very clearly in table 7.2 that very high levels of communicative success are always achieved under these circumstances, with almost negligible levels of variation after 5000 communicative episodes. It does appear, however, that these very high levels of communicative success are slightly less likely to occur as comprehensive depth increases. Increasing the comprehensive depth, of course, gives the agents many more potential meanings to decipher, and so it is perhaps not altogether surprising that it takes them longer to work out the meanings of all the words, some of which will of course be used only very rarely. In all cases, however, the κ rate tends towards 100% with a graph essentially that shown in figure 7.3, with slight differences in the speed at which the initial spurt occurs, and the rate of increase of the curve in the latter part of the experiment;

[2]The standard deviation is scaled relative to the mean so that we can more accurately compare results from distributions with different means.

we can safely say that, under these experimental conditions, the introspective obverter method can indeed develop communication systems with a very high success rate.

Even though the actual target object referred to by the speaker is not known, the hearer does have access to the context in which the signal was uttered. The disposition to only consider whole objects as possible referents in accordance with Macnamara (1982)'s whole object bias, discussed in section 3.3.2, which we are assuming in the model, reduces the uncertainty of reference to a finite, rather than the infinite problem with which Quine (1960) was grappling. This finite problem, although soluble, can still be very large, as each utterance can be paired with all meanings which could discriminate any one of the objects in the context from all the other objects in the context; even for each object there could be multiple semantic hypotheses, for instance an object could be described as TURQUOISE, EXTREMELY LONG or CIRCULAR, all of which might in principle identify it from the other objects in any particular communicative episode. The context in which the signal was uttered, as we saw in section 6.3.3, provides a form of restricted joint attention which allows the hearer to solve the uncertainty problem practically; in section 6.5, we saw that the agents maintain a record of all the possible meanings which have ever been associated with a signal, through the confidence probability which is stored with each signal-meaning pair the agent has encountered. After the signal has been uttered many times, it will have been paired with many possible meanings, and the agent will therefore have considered, and have a record of, many different semantic hypotheses. Many of these meanings will have been encountered in more than one context, and the agent's confidence in these particular signal-meaning pairs will rise as a consequence. Over time, one meaning will occur in more contexts than any other, and that meaning will be considered by the agent as the most likely meaning to be associated with the signal.

7.3.2 Non-Synchronised Comprehensive Conceptual Structure

In this section, I present similar experiments to those we have just looked at, in which the agents are once more provided with comprehensive innate conceptual structure, but in these simulations the agents' meaning structures are not synchronised. There are a number of ways of implementing this; one of the most straightforward and interesting is to limit the number of channels on which innate meaning structure is provided. One agent, for instance, might have a comprehensive conceptual structure on all five of its channels, but the other agent only has meaning structure on four of its channels; although every channel which has a discrimination tree is refined in the same comprehensive manner and to the same depth, as shown in figure 7.1, the fifth channel is left without conceptual structure at all. In this particular case, setting up one channel out of five without meaning structure would result in a pre-defined meaning similarity rate $\sigma = 0.8$ between the two agents. This method of conceptual structure provision presents an easy way to set up experiments so that the two agents have a specified amount of meaning similarity σ; table 7.3 and figure 7.4 both show results from experiments in which both agents have comprehensive structure to depth 3, but whereas the speaker has this structure on all five of its sensory channels, the hearer only has structure on a specified number of channels, so enabling the meaning similarity between them to be fixed at regular intervals between 0 and 1. Again, each of the experiments was run 50 times at each level of meaning similarity, and the summary statistics were then calculated. We can see clearly that there is a strong relationship between meaning similarity σ and communicative success κ under these conditions. When $\sigma = 1$, as we have already seen in the previous section, the mean communicative success $\bar{\kappa}$ is also very high, approaching the level of σ, and there is very little variation. As σ decreases, $\bar{\kappa}$ also falls, but not as as quickly as σ; at $\sigma = 0.8$ they are at approximately the same level, but when $\sigma = 0.2$, both the mean $\bar{\kappa}$ and indeed the minimum level of κ seen are considerably higher than the meaning similarity level. Moreover, the variation in κ increases considerably as σ decreases, with $\mathrm{CoV}(\kappa)$ increasing from 0.01 to 0.16. There is, therefore, a clear relationship between the level of innate meaning similarity and the level of communicative success which the agents are likely to achieve: as σ increases, the level of communicative success κ increases, and the variation in the likely level of κ decreases considerably.

When the hearer is missing conceptual structure on one or more of its sensory channels, then the introspective obverter algorithm can only produce levels of κ which are closely related to the level of meaning structure σ. When meaning similarity is high, identity of both sense and reference will eventually occur through disambiguation, and so the levels of κ and σ are closely related, but when meaning similarity is low, even though sense identity is often impossible, reference identity can still occur, leading to relatively higher rates of κ compared to σ. For instance, if the speaker utters a signal corresponding to a meaning on the particular sensory channel which the hearer lacks, then there is no possibility that disambiguation over multiple contexts will ever lead the hearer to the speaker's meaning, because the hearer can never consider a semantic hypothesis to which it does not have access. Paradoxically, however, because of the hearer's conceptual deficit, it will actually have fewer semantic hypotheses to consider for the signal, and may therefore be able to settle on a meaning more quickly than if it had many possible meanings to disambiguate. The hearer's meaning will not be that which was intended by the speaker, but because communicative success is based on reference identity and not on sense identity, the meaning may still refer to the appropriate object.

Meaning	Communicative Success Rate				
Similarity σ	$\bar{\kappa}$	(CI)	Max(κ)	Min(κ)	CoV(κ)
1.00	0.96	(0.96 – 0.97)	0.98	0.91	0.01
0.80	0.82	(0.80 – 0.84)	0.92	0.70	0.08
0.60	0.66	(0.64 – 0.68)	0.78	0.51	0.11
0.40	0.52	(0.50 – 0.54)	0.63	0.38	0.12
0.20	0.34	(0.32 – 0.36)	0.50	0.24	0.16

Table 7.3: Meaning similarity σ and mean communicative success $\bar{\kappa}$. Agents are innately provided with meaning structure at various levels of meaning similarity σ, and then undertake 5000 communicative episodes. Each simulation is repeated 50 times, over which the mean communicative success rate $\bar{\kappa}$, together with a 95% confidence interval around the mean, the range of values for κ, and the coefficient of variation, are calculated.

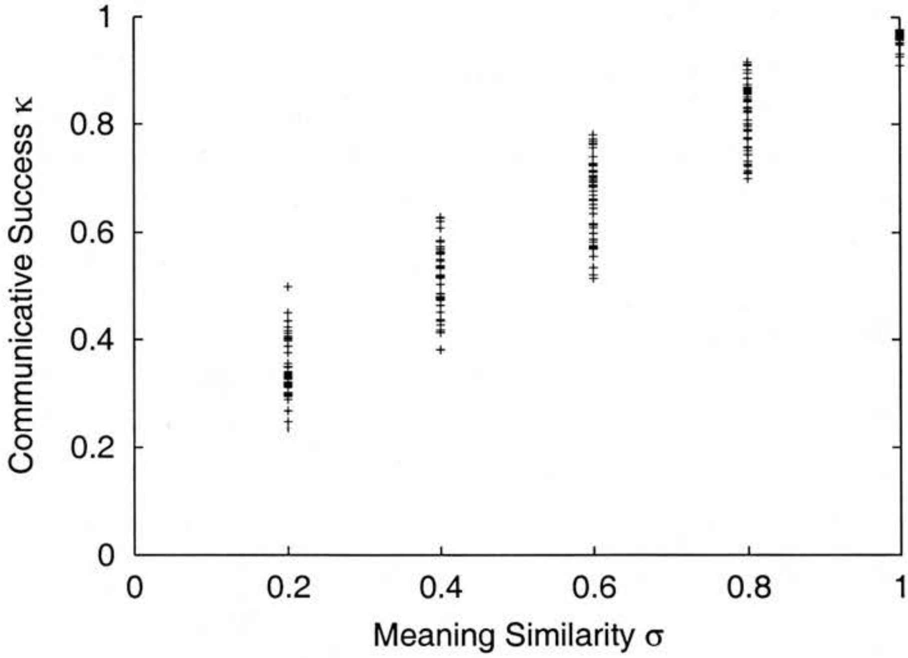


Figure 7.4: Meaning similarity σ and communicative success κ. Agents are each provided with comprehensive innate conceptual structures with some channels restricted to produce various fixed levels of meaning similarity σ. Meaning similarity is plotted against the level of communicative success after 5000 communicative episodes. The experiment was repeated 50 times at each level of meaning similarity, with each cross on the graph representing the results from a single run.

Variable	Mean	(CI)	Max	Min	CoV
κ	0.93	(0.92 – 0.94)	0.97	0.88	0.02
σ	0.66	(0.65 – 0.67)	0.73	0.60	0.05

Table 7.4: Communicative success κ and meaning similarity σ in agents with innate random conceptual structures. A summary of the range and distribution of κ and σ after 5000 episodes, averaged across the 50 runs of the experiment shown in figure 7.5

7.3.3 Random Conceptual Structure

All these experiments, of course, were carried out after providing the agents with *comprehensive* innate conceptual structure; what happens if the agents are instead provided with random conceptual structure? With comprehensive innate structure, it is possible to determine in advance the exact level of meaning similarity that the agents in the experiment would have, and, in the experiments described above, I let this range from synchronised ($\sigma = 1$) to very dissimilar meaning structures ($\sigma = 0.2$). With random conceptual structure, on the other hand, the level of meaning similarity which emerges is not deterministic at all, and could in principle vary quite dramatically between runs of the same experiment; one agent might develop a particular sensory channel to a very great degree, allowing a very sensitive categorisation of objects in respect of their colour, for instance, while the other agent might have only a very coarse representation of colour, but can instead differentiation many objects by their smell.

In this section, therefore, we will look at experiments in which most settings remain the same, but the agents are each provided with 60 random innate concepts, or nodes on their discrimination trees, as shown in figure 7.2, before they attempt to communicate with each other. As before, each simulation is run 50 times, with each run being shown as a separate line in figure 7.5, and, after 5000 communicative episodes, summary statistics are calculated and displayed in table 7.4. Because of the random setup of these experiments, both communicative success κ and meaning similarity σ vary over the 50 runs, and so 7.4 shows summaries for both variables. Figure 7.6 shows a more detailed breakdown of the relationship between meaning similarity and communicative success, with a plot of the two variables against each other, after 5000 communicative episodes have been carried out. We can see immediately in table 7.4 that the mean level of communicative success $\bar{\kappa}$ is high, and, as we saw previously, there is little variation in κ under these circumstances. Perhaps surprisingly, however, the mean level of meaning similarity $\bar{\sigma}$ is much lower, at only 66%, and again with little variation. In figure 7.6, we can confirm

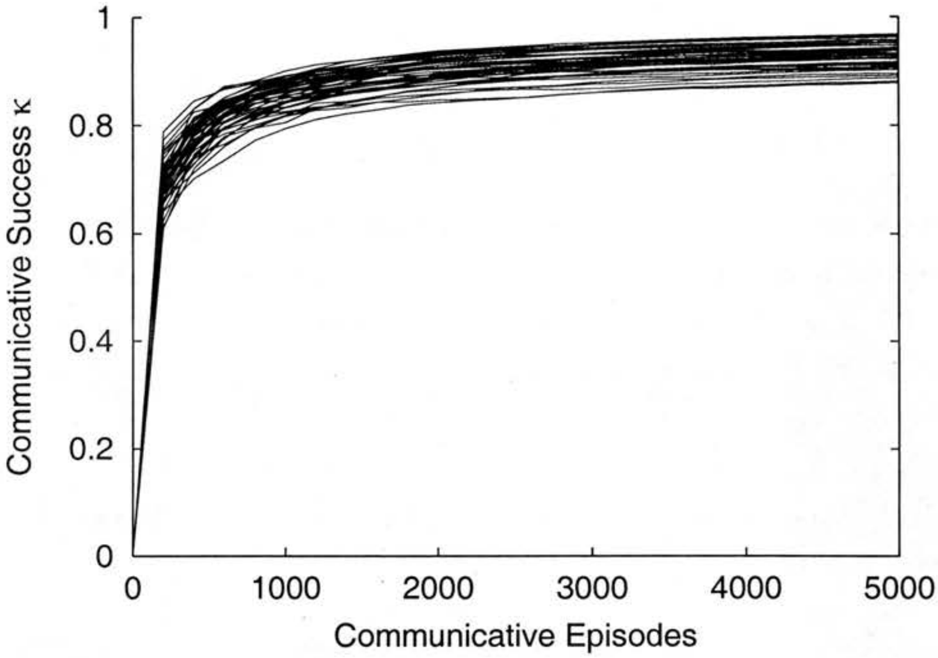


Figure 7.5: Communicative success κ in agents with innate random conceptual structures. Agents are each provided with 60 random innate meanings, and the experiment was repeated 50 times, with each line on the graph representing a single run.

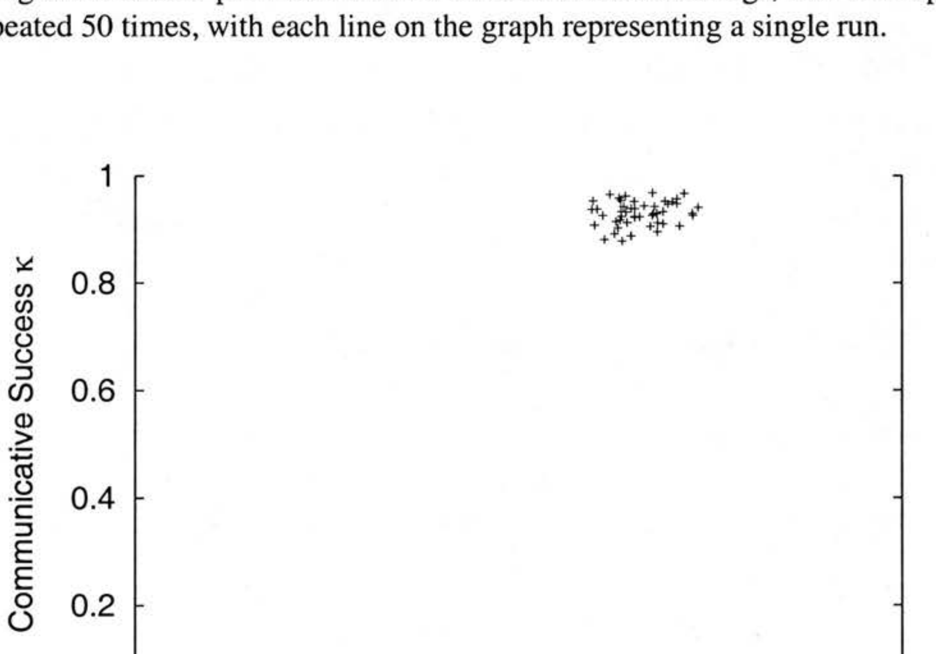


Figure 7.6: Communicative success κ and meaning similarity σ after 5000 communicative episodes, in agents with innate random conceptual structures. Each cross on the graph shows the endpoint of one run of the simulation shown in figure 7.5.

that neither variable shows much variation at all, as the plots on the graph are clustered together very tightly.

7.3.4 Summary

The contrasting communicative success results we find when different types and levels of innate conceptual structures are provided to the agents present us with an intriguing puzzle. When the agents have comprehensive structure, there is a clear linear relationship between the level of meaning similarity σ and the level of communicative success κ; as σ increases, κ also increases, albeit slightly more slowly. Moreover, the variation in κ decreases quite substantially as κ approaches 1 due to a ceiling effect, showing that agents with identical meaning structures will communicate optimally using introspective obverter.

When the agents' structure is randomly generated, on the other hand, the relationship between meaning similarity and communicative success is more opaque; there is much less variation in the level of σ than we might have expected, and therefore insufficient information to confirm the slight upward trend in κ which appears to be present, for instance in figure 7.6. Moreover, the levels of communicative success achieved are considerably higher than those we might have expected from our results with innate structure shown in figure 7.4. What are the crucial factors concerning the production and use of random conceptual structure which results in such elevated levels of communicative success?

7.4 Randomness and Predictability

When we are experimenting with comprehensive innate conceptual structure, the crucial parameter is, as we have seen, the level of meaning similarity between hearer and speaker. But with random conceptual structure, we find much less variation in both meaning similarity and in communicative success; moreover, the levels of communicative success are higher than we would expect given the comprehensive results. In order to explain this, we must delve deeper into the process of random meaning provision, and in particular we need to differentiate two different kinds of randomness in the allocation of random innate meaning structure, whose effects are subtly different, as follows:

- a sensory channel is chosen at random;
- a value on the selected channel is then chosen at random, and the leaf node λ corresponding to this value is refined.

The random choice of a sensory channel is very straightforward, and need not concern us very much; firstly, the fact that there is a uniform probability distribution across the channels implies that each channel has a probability $\frac{1}{n}$ of being chosen, where n is the number of channels; secondly, the relatively large number of meanings chosen (60 in this case) for each agent, means that the distribution of channels chosen will be very similar between agents.

Once the sensory channel has been chosen, however, the choice of node which is refined is crucial in determining the similarity τ (see equation 5.13), and therefore indirectly the meaning similarity σ between agents. Furthermore, the fact that our meanings are created on binary trees leads to some interesting (and perhaps unforeseen) consequences. If, for instance, a channel is refined once, then we know exactly what the tree will look like after the refinement; there is only one way in which an empty tree can be refined, and it always produces a tree with two leaf nodes, shown as A at the top of figure 7.7.

Moving down this diagram, we follow the stages through which a discrimination tree passes as it is randomly refined; if the tree in A is refined for a second time, it is clear that either of its two leaf nodes could be chosen, each with an equal probability of 0.5, as shown by the annotations to the dotted lines, giving two possible trees B and C, each of which has three leaf nodes[3]. If a third refinement takes place, any of the three leaf nodes on B or C can be refined, but because of the hierarchical way in which subsequent levels of the tree divide up the semantic space, it is important to note that each node is not equally probable; the node at depth 1 has a probability of 0.5 of being chosen, while the two nodes at depth 2 each have a probability of 0.25. In fact, because of the binary nature of the trees, the probability of a particular leaf node being chosen is given simply by $\frac{1}{2^d}$, where d specifies the depth of the node. In order to work out the probability of any particular path being traversed, we simply multiply the probabilities we find on the path back from that tree to the tree at the top of the diagram[4].

Calculating the probability of any particular *tree*'s occurrence is likewise straightforward, and is simply the sum of the probabilities of each path which could lead to that tree; the probabilities for the trees D-H, which are obtained after three refinements, are shown in brackets at the bottom of figure 7.7. Interestingly, although there are five possible trees at this depth level, there are marked differences in the chances of obtaining each of them. Tree F, in particular, can be reached from both trees on the preceding level, B and C,

[3]Every refinement of a tree, as we can see, increases the number of leaf nodes which are available for future refinement by one node.

[4]More properly, we would need to multiply the resultant probability by the probability of obtaining this first tree, but we have already seen that this latter probability is 1, and so the term can be eliminated from the equation.

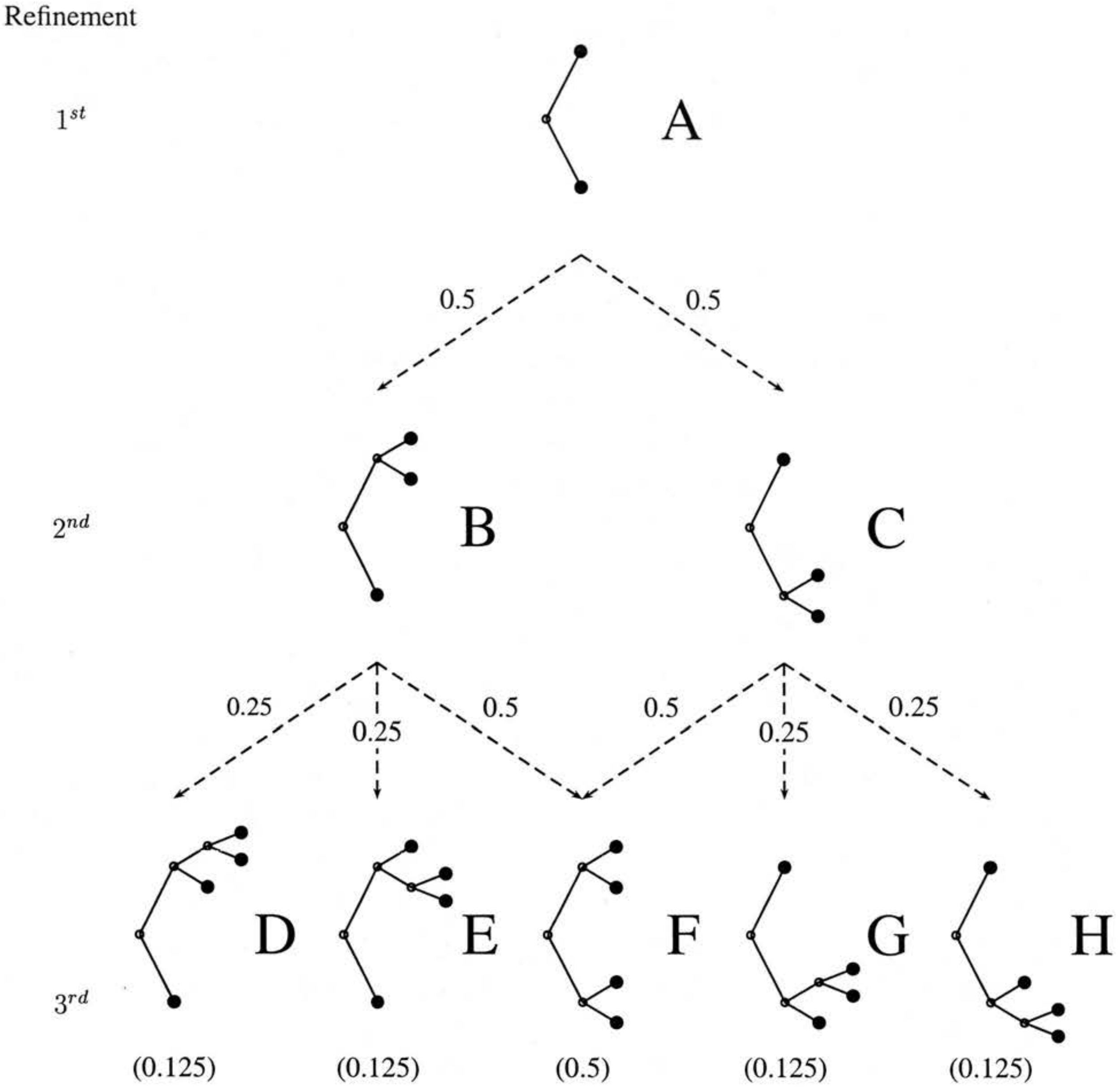


Figure 7.7: The random refinement of a sensory channel. The first refinement of a sensory channel is deterministic, and always results in the tree at the top of the figure (A). From this tree, either of the leaf nodes (the slightly larger, black nodes) can be chosen, with a probability of 0.5 (shown on the dotted lines leading from tree A), resulting in the trees on the second level (B and C). Thereafter, the number of possible trees increases dramatically (see equation 7.2), but the available nodes are not equally probable, and this pressure leads to a trend for the trees to become as comprehensive as possible (compare tree F to the other trees on the lowest level).

Tree	D	E	F	G	H
D	1	0.67	0.67	0.33	0.33
E	0.67	1	0.67	0.33	0.33
F	0.67	0.67	1	0.67	0.67
G	0.33	0.33	0.67	1	0.67
H	0.33	0.33	0.67	0.67	1

Table 7.5: Levels of mutual tree similarity τ between trees D-H in figure 7.7.

whereas trees D, E, G and H each have only one path which reaches them. If we focus in more closely on either tree B or tree C, we can also see that the path which leads from here to tree F is twice as probable as either of the paths which lead from this tree to the other successor trees. Tree F, in fact, the only comprehensively refined tree of the five, will occur fully 50% of the time, while each of the others will occur just once in every eight attempts. The random nature of the meaning allocation process, therefore, coupled with the hierarchical nature of the semantic structure, combine together to create a considerable pressure towards comprehensively refined trees.

Tree F in figure 7.7 is clearly much more likely to occur than D, E, G or H, but what effect might this have on meaning similarity? Table 7.5 shows a matrix of tree similarity τ for all combinations of trees D-H, and even with just five possible trees under consideration, we can find a number of interesting properties. Firstly, the lowest levels of τ (0.33) are found when one of DE (the trees which can only be reached from B) is compared with one of GH (the trees which can only be reached from C); it is clear that low levels of τ therefore can only occur when the meaning structures diverge from an early stage in the meaning creation process, and that both trees being compared are not comprehensive, but instead are very *specialised*, in that they can make very fine distinctions at one part of the feature spectrum, but not anywhere else. In contrast, apart from the obvious identity relation when a tree is compared with itself, the highest values found anywhere in the matrix are when tree F is compared with any other tree, which is always the same (0.67). If we couple this with the knowledge that tree F is also much more likely to occur than any of the other trees, what will this mean for an expected level of tree similarity $exp(\tau)$ between a random combination of any two trees created with three refinements? The general solution to this is given by the following equation:

$$exp(\tau) = \sum_{i=0}^{n} \sum_{j=0}^{n} \tau(t_i, t_j) p(t_i) p(t_j), \tag{7.1}$$

where n is the number of possible trees under consideration, and $p(t)$ is the probability of tree t being chosen. For the five trees D-H in figure 7.7, therefore, $exp(\tau) = 0.729$, which is perhaps higher than we might have imagined, given that this occurs through a completely random process of concept creation. As we have seen, however, it is the random nature of the concept creation process itself which leads to pressure to create comprehensively refined trees. Comprehensively refined trees in turn produce the highest possible levels of tree similarity τ when compared with trees with the same number of leaf nodes (with the exception of comparing a tree with itself), leading to expected levels of τ far above what we might initially imagine.

It is arguable, however, that such high levels of τ are an artefact of the situation with trees which have only been refined three times; little meaning creation has happened, and this inevitably means little variation. If we therefore increase the number of refinements to four, the number of different possible trees which can be created rises to 14; looking at these possible trees in detail, we can see that:

- eight of them are the most specialised trees, which can only be reached by refining one of the deepest nodes on trees D, E, G or H (figure 7.7). Each of these occurs with a probability of just 0.016;
- two of them are reached by refining the node at depth 2 on trees D, E, G or H. Each of these trees has two different routes to creation, and occurs with a probability of 0.063;
- the final four trees can be reached either by refining any of the nodes on the comprehensively refined tree F, or by refining the node at depth 1 on any of the other trees. There are therefore two ways to reach all of these trees, and again the particular paths needed to reach these are relatively more likely, so each of these four occurs with a probability of 0.188.

Not surprisingly, the tree similarity levels are also highest for these latter four trees, which are by far the most likely to occur, with a combined probability of 0.75. Although there cannot be any completely comprehensive trees created from four refinements, I hope that it is clear that this pressure to create trees which are as balanced as possible is still very strong. As we increase the number of refinements n, the number of different possible

Refinements n	Possible Trees $\phi(n)$	Mean $\bar{\tau}$	(CI)	Range Max(τ)	Min(τ)	CoV(τ)
3	5	0.740	(0.700 – 0.780)	1.000	0.333	0.275
4	14	0.745	(0.713 – 0.777)	1.000	0.250	0.217
5	42	0.692	(0.659 – 0.725)	1.000	0.200	0.243
6	132	0.698	(0.670 – 0.726)	1.000	0.333	0.202
7	429	0.709	(0.683 – 0.734)	1.000	0.429	0.180
8	1430	0.693	(0.671 – 0.714)	0.875	0.250	0.160
9	4862	0.688	(0.668 – 0.708)	1.000	0.444	0.146
10	16796	0.684	(0.663 – 0.705)	0.900	0.400	0.158
11	58786	0.695	(0.675 – 0.716)	0.909	0.455	0.152
12	208012	0.683	(0.661 – 0.705)	0.917	0.333	0.162

Table 7.6: Observed levels of tree similarity τ between randomly created trees with different numbers of refinements. One hundred separate pairs of trees were created, and a summary of the mean and range of τ over these experiments is shown.

trees ϕ_n increases dramatically, following the series of Catalan numbers, each member of which is given by the equation

$$\phi_n = \frac{(2n)!}{(n+1)!n!} \tag{7.2}$$

(Conway & Guy, 1996), leading rapidly to very intense and laborious calculations to discover the expected level of tree similarity for all possible combinations of these trees. To avoid the need for these calculations, I have instead run computational simulations, in which two trees are created at random with a certain number of refinements, and the tree similarity level then calculated explicitly. Each experiment is run 100 times, and the average observed tree similarity level, together with other summary statistics, is displayed in table 7.6, and should provide a good approximation to the mathematically expected rates.

We can see in this table that the experimental results are indeed consistent with the theoretical results which we had anticipated. The expected value of τ at $n = 3$, for instance, derived from equation 7.1, of 0.729, falls well within the confidence limit (0.70-0.78) found experimentally. Even more interestingly, the pressure to keep trees balanced and as comprehensive as possible results in remarkably similar levels of tree similarity, no matter how many nodes are created on the trees. There is a very small decline in the level of $\bar{\tau}$ as we move down table 7.6 and the tree structures become ever more complicated, coupled however with a reduction in the variation $CoV(\tau)$. Random concept allocation, therefore produces structures which are progressively more similar to each other as more conceptual structure is created.

Finally, because the expected level of τ in two randomly created trees is relatively static, it is not surprising to find that the corresponding level of σ, the meaning similarity between agents, which is averaged across all their sensory channels, also had little variation in the experiments which we looked at in figure 7.6.

7.5 Semantic Generality

But why is the level of average communicative success $\bar{\kappa}$ in table 7.4 so much higher than we might have expected, had we taken the experiments with non-synchronised comprehensive meanings as faithfully indicative of the relationship between meaning similarity and communicative success? In order to answer this, we need to focus on the communicative process in the context of Grice (1975)'s philosophical model of conversation. In the Gricean model, it is proposed that the communicative process is governed by the cooperative principle, a set of hypothetical, implicit, rules which underlie communication. Famously, Grice unpacked this principle into four conversational maxims, reproduced below:

Maxim of Quantity: Be informative.

Maxim of Quality: Be truthful.

Maxim of Relation: Be relevant.

Maxim of Manner: Be perspicuous.

Of course, these maxims are often violated in conversation, for rhetorical effect such as irony or sarcasm, and their violation allows the the hearer to make certain conversational implicatures. Implicatures, in turn, allow the construction of additional aspects of meaning which are not explicitly referred to in an utterance. My model of communication between agents is of course not meant to be a complete account of communication, and it assumes indeed that the Gricean maxims are not violated. Of most relevance in this discussion is the first maxim above, that of quantity, found in the epigram at the start of this chapter and in detail below:

- Make your contribution as informative as required.
- Do not make your contribution more informative than required.

In my model of communication, the agents use meanings which provide sufficient information to identify the target object, but not unnecessarily specific information. To take a hypothetical example, if a speaker has a very detailed discrimination tree, for instance, on the channel which represents colour, yet in a particular discrimination game the target object can be described as RED, while all the other objects are BLUE, the agent will use the general meaning RED rather than a very specific meaning DEEP VERMILION. The purpose of communication in this model, after all, is to identify one particular object from a group of other objects, and in this context the use of the specific meaning DEEP VERMILION would be considered as providing more information than was required. Likewise, when interpreting an utterance, the hearer will conform to the same maxim of quantity when constructing its list of semantic hypotheses, discarding a possible meaning DEEP VERMILION in favour of RED, as long as the meaning RED is sufficiently specific to identify one object in the context.

In the hierarchical dendritic semantic model in use in these experiments, it is clear that general meanings can be defined as those which are nearer the root of the discrimination tree, and specific meanings are those nearer the leaves of the tree. We saw in section 7.3.2 that it is important from a communicative point of view for the hearer to contain in its conceptual structure those meanings which the speaker is likely to use in the communicative episodes. The Gricean preference for the use of more general meanings over specific meanings by both agents means that these general meanings are those which the hearer needs to have in its semantic repertoire for communication to succeed. In section 7.4, we noted that, coincidentally, there is a strong bias in the random meaning creation process for balanced, comprehensive discrimination trees, and consequently for a full range of general meanings to occur in the agents' trees. Differences between agents' conceptual structure, therefore, are more likely to occur towards the leaves of the discrimination trees, in the parts of the semantic structure which are less important for communicative success.

To summarise, the Gricean maxims of communication in the model result in general meanings being disproportionately used by the agents, while the hierarchical structure of the meaning structure and the random nature of the meaning creation process result in general meanings being disproportionately shared by the agents. Together, these lead to a communicative success rate which is considerably higher than the level of meaning similarity (see table 7.4) alone. Meaning similarity remains a very important predictive factor for communicative success rates when we are relying on the disambiguation of utterances through context, but it will always, except in the special case when $\sigma = 1$, underestimate the actual rate of communicative success due to the Gricean preference for objectively more general meanings over specific ones.

7.6 Summary

In this chapter, I have investigated the efficacy of the communicative framework based on my introspective obverter algorithm, in which the hearer is told neither the speaker's intended meaning, nor the referent of this meaning, nor given any feedback concerning the success of its attempted interpretation. In section 7.3, I broke the link between the process of unguided meaning creation described in chapter 5 from the process of unguided communication described in chapter 6, and provided agents with innate meanings in various configurations. I showed that, given innate conceptual structure under this variety of circumstances, the communicative success rate κ is consistently very high:

- If agents have identical comprehensive conceptual structures ($\sigma = 1$), then communicative success is near-perfect (table 7.2);
- If agents have comprehensive structure which is not synchronised, because the hearer is lacking a certain number of sensory channels, then communicative success κ is very strongly correlated with meaning similarity σ (figure 7.4);
- If agents have randomly allocated conceptual structure, then σ is regularly between 65-70%, but communicative success is consistently much higher at around 93% (table 7.4), due to the Gricean preference for general meanings discussed in section 7.5.

In section 7.3.3, I explored the provision of random meanings, and found that communicative success is in this case always at a higher level than the meaning similarity. I then investigated the reasons for this both theoretically and experimentally, in section 7.4, and made the important discovery that:

- the random nature of the meaning allocation process, coupled with the hierarchical nature of the semantic structure in the discrimination tree mode, combine to exert considerable pressure in favour of the construction of balanced, comprehensive meaning structures at similar levels of tree similarity.

I then went on to demonstrate in section 7.5 that the meanings which are most likely to be shared in these structures, namely the more general ones, are also most likely to be used by the agents in accordance with Gricean conversational principles. In chapter 8, I will go on to explore experimental conditions under which levels of meaning similarity are not as uniform and predictable as in the models described here, and show how agents build highly co-ordinated structures despite the pressures inside the system.

CHAPTER 8

Meaning Creation and Communication

> "[S]imple heuristics perform well . . . if the structure of the heuristic is adapted to that of the environment." (Gigerenzer & Todd, 1999, p. 13)

8.1 Introduction

I have identified a number of characteristics of the meaning creation process as it stands so far which help to explain the mechanisms and conditions which help to facilitate communication using the introspective obverter algorithm. In particular, we saw in section 7.4 how the random nature of the meaning creation process creates a strong pressure to develop comprehensively refined trees, and indirectly leads to fairly predictable levels of meaning similarity between the agents, though this predictability decreases slowly as we increase the number of channels on which an agent can grow conceptual structure. The agents can, furthermore, develop a co-ordinated system of communication by inferring an utterance's meaning through context-driven disambiguation, despite having no access to each other's internal lexicons and no guidance about what or how well they are learning the meanings of the vocabulary items.

In this chapter, my investigations will proceed through comprehensive computational experiments based on the model I have described, and will focus in detail on two particular areas:

- the re-linkage of meaning creation and communication, so that agents create their own grounded meanings, rather than having them provided by the model.

- the exploration of how various cognitive biases and environmental structure influence the levels of meaning similarity and communicative success.

Firstly, the re-linkage of meaning creation with communication will show us whether agents can communicate as effectively with their own, individually created meanings, as they could with the meanings of various kinds provided by the model in chapter 7. The experiments, therefore, are reconfigured so that both agents develop their own individual meaning structures, as discussed in chapter 5, by responding to failures in their attempts to discriminate certain objects from other objects in their external environment. In a standard simulation, each agent plays an average of 1000 discrimination games, and thereby develops a semantic structure which successfully represents the world around them. In each discrimination game, the agent tries to discriminate one particular object from a larger context of five objects; if this is not possible, then the agent chooses a sensory channel, and refines the node which describes the target object; for instance, the agent might choose the *colour* channel[1]. If the target object was blue (but at least one other object in the context is also blue, hence the failure of the game), then the agent could refine the category BLUE (which describes the target object) into the more specific categories DARK BLUE and LIGHT BLUE. Having individually created their meaning structures, the agents will play 5000 communicative episodes as described in chapter 6, with one agent acting as the speaker, and the other agent acting as the hearer. As before, communicative success is defined in terms of *referent* identity (see section 6.3.2), and the hearer is provided only with the signal and the context in which the speaker uttered it; neither agent, of course, receives any information about the success of the communicative episode.

Secondly, the exploration of the model under a number of different circumstances and conditions will not only allow me to test the robustness of the theory of communication through the inference of meaning, but will also provide evidence concerning the utility of the cognitive and environmental factors under investigation. In section 8.2, I will explain how cognitive biases are introduced into the model, and explore how they interact with different meaning creation algorithms to produce varying levels of meaning similarity and communicative success in the standard world we have looked at so far. I will then move on in section 8.3 to modify the simulations so that both agents have the same experiences of the world they inhabit, and we will see that this can, in certain circumstances, lead the agents to produce completely synchronised conceptual structures, and therefore optimal communication systems. In section 8.4 I will investigate modifying the structure of the environment itself, developing a more realistic world for the agents to inhabit,

[1] Remember that the sensory channels in these experiments are in fact totally abstract, and names such as *colour* are only used for expository purposes.

and I will show that, when this is in place, the agents can develop extremely successful communicative systems merely through the inference of meaning, as they exploit the structure of the environment.

8.1.1 Experimental Measurements

In the body of this chapter, I will present many experimental results which show variations in meaning similarity σ and communicative success κ across various different parameters, which will themselves be introduced in the following sections. As a general rule, each simulation is repeated 50 times, over which the now-familiar summary statistics for the levels of both σ and κ are calculated.

In addition to the raw results, it is interesting to compare the distribution of results obtained by one set of experiments to those obtained by another set, to see if there is any statistically significant difference between them. In order to measure this, I use the Kolmogorov-Smirnov (KS) statistic to express how different two sample distributions are. Strictly speaking, we are trying to disprove the null hypothesis that the two datasets are drawn from the same population distribution function (Conover, 1999).

In the tables in the forthcoming sections, the notation KS($x.y$) represents a statistical comparison between the data in the current (row of the) table and the data in (the corresponding row of the) table $x.y$. In general, the higher the level of the KS statistic, the more certain we are that the two distributions are not from the same underlying population; crucial to the interpretation of the Kolmogorov-Smirnov statistic is its significance level, *small* values of which show that the distributions are indeed different. Significance levels under 0.05 will be denoted by an asterisk (*), and those under 0.01 will be indicated by a double asterisk (**).

Although the most interesting and relevant results are presented throughout this and the following chapter, I have also produced full details of the results for all the experiments in appendix C for reference.

8.2 Cognitive Biases and Tree Growth Strategies

I have already discussed, primarily in chapter 3, many proposed solutions to the apparent paradox of how children manage to acquire their lexicon, and, in particular, that scientists have appealed to many different cognitive biases to solve this problem, some of which I will briefly recap here:

Whole-Object Bias (Macnamara, 1972): a child will assume that an unfamiliar word names a whole object, rather than a particular property of it;

Shape Bias (Landau et al., 1988): a child is more likely to assume that an unfamiliar word refers to the shape of an object rather than to other properties such as its colour or taste;

Taxonomic Bias (Markman & Hutchinson, 1984): a child will group the same kinds of objects together;

Underlying all these proposals are variants of the idea that learners are pre-disposed to focus on particular properties of objects they are learning names for, and it is easy to see analogies with the sensory channels in this model of meaning creation. In this section, and throughout the analyses in this chapter, therefore, I will investigate how abstract biases affect the construction of conceptual categories. However, I do not propose to investigate whether, for instance, the shape bias accounts for more of the acquisition process than the taxonomic bias, because I have designed the model to be as abstract as possible on purpose, and the channels to be intrinsically meaningless, so it makes no sense to arbitrarily decide that channel one should be renamed 'shape' and channel two 'smell', and then investigate which is better for the agents. Instead, I will investigate cognitive biases in the abstract, focusing on what kind of biases are important for communication to succeed. In particular, implicit in all the explanations of lexicon acquisition we have looked at are that the biases and assumptions are 'sensible', and made by *all* children, that is, they are universal biases. I will investigate in this section whether it is indeed important for communication that agents have identical cognitive biases, or whether they can communicate successfully despite not having the same cognitive biases.

8.2.1 Sensory Channel Biases

In the simulated world within the model, I have already described how each agent a is created with a set of numbered sensory channels, which correspond to the features by which the objects in the world are defined. The evolution of these channels is not under consideration here, though this of course remains an important question which is the subject of much contemporary research (see Polani, Uthmann, and Dautenhahn (2000), Ward, Gobet, and Kendall (2001), Ziegler and Banzhaf (2001), among others). Each sensory channel c has a bias b_{a_c}, which is stored as a real number in the range $0 \ldots 1$, and can be thought of as representing the probability of the channel being chosen for refinement[2]. The channel biases do not change over time, so the method used to set them in the initialisation of the experiments is crucial. In the standard model we have been using hitherto, my exposition has ignored the channels biases so far, because I have set them so that, for each agent, every channel bias is equal (i.e. there is a uniform bias distribution across the agent); the agent essentially chooses a channel at random each time a channel is needed. The channel biases, however, can of course be defined according to many different probability distributions; I will investigate three particular interesting cases in detail, including the uniform distribution used in all the experiments described in chapter 7:

Bias Allocations

Uniform Bias Allocation, in which the agent's channel biases are all equal;

Random Bias Allocation, in which each of the agent's channel biases is a randomly generated number in the range 0.0 ... 1.0;

Proportional Bias Allocation, in which each of the agent's channels has a bias which represents a fixed proportion p of the remaining bias available to the agent, taking into account biases which have already been allocated to its other sensory channels. This description might seem a little opaque, so let us first consider that the biases form a probability distribution, and therefore the total value of all the biases for an agent must equal 1. The remaining bias available is defined as the total of all biases on the channels whose biases have already been set, subtracted from 1. With this in mind, b_{a_c} is defined as in the following equation:

$$\begin{aligned} \text{if} \quad c = 0, \quad b_{a_c} &= p \\ \text{if} \quad c > 0, \quad b_{a_c} &= p\left(1 - \sum_{i=a_0}^{a_{c-1}} b_{a_i}\right), \end{aligned} \tag{8.1}$$

Because the biases represent a probability distribution, they must always be scaled, no matter what allocation method is used, so that the following equation always holds for each agent:

$$\sum_{i=0}^{c-1} b_{a_i} = 1,$$

[2]We will see soon that such a representation actually only makes sense under the probabilistic tree growth strategy, but it is a useful starting point.

Channel	Bias	Scaled Bias
0	0.5	0.5161
1	0.25	0.2581
2	0.125	0.129
3	0.0625	0.0645
4	0.03125	0.0323

Table 8.1: Allocation of biases under the fixed proportional method, with $p = 0.5$.

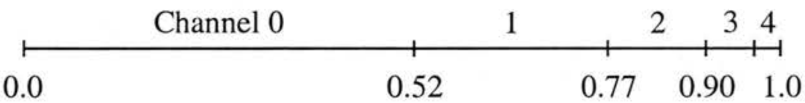


Figure 8.1: A 'bias line' representing the probability distribution of sensory channels, whose biases are proportionally ($p = 0.5$) created.

namely that the sum of the channel biases for the agent always adds to 1. For instance, under proportional bias allocation, if $p = 0.5$, and there were five sensory channels, the biases would be allocated as shown in table 8.1. Another way of thinking of the biases as a probability distribution is to visualise them as in figure 8.1, which shows a line from 0 and 1 representing the probability distribution across all of an agent's channels. A random number r in the range 0 ... 1 is chosen, shown below the bias line, and the corresponding channel, shown above the line, is chosen for refinement and the creation of new meaning structure. It is worth observing here that the proportional method of channel selection is deterministic, so if two agents have the same value of p, then they will necessarily have identical cognitive biases[3].

8.2.2 Tree Growth Strategies

As well as changing the distribution of an agent's channel biases, we can also investigate variations in the strategies the agents use for channel selection, which I call *tree growth strategies*. The tree growth strategy determines the method of sensory channel selection, but not the particular leaf node λ on the tree which will be refined; in all cases this node is defined as follows:

leaf node to be refined $\lambda(c)$: the deepest node in the discrimination tree on the sensory channel c which categorises the target object in the current discrimination game.

[3]Unless specified otherwise, p is set to 0.5 for all simulations reported in this thesis.

Firstly, in the default or *probabilistic* method, the agent chooses a sensory channel on the basis of the channel biases, as described in section 8.2.1 and shown in figure 8.1. In addition to this method, I will investigate another strategy, in which the agent first orders its channels according to their biases, then searches through them, considering possible nodes λ which would have resulted in successful discrimination of the target object *in this particular discrimination game*, had λ already been refined. If no channel which meets this criterion is found, then no refinement takes place. Three noteworthy features of the second strategy, which I will call the *intelligent tree growth strategy*, stand out immediately (others will become apparent throughout the course of this chapter):

- an intelligent refinement will always make a helpful distinction in at least the particular discrimination game during which it was created. By contrast, refinements under the probabilistic strategy are not guaranteed to be successful in any future discrimination games at all.
- if there are no channels on which a possible refinement would have been successful in the current discrimination game, then no refinement takes place at all. Such a situation is more likely to arise at the beginning of a simulation, before trees of any great depth have been grown, so it seems reasonable to assume that the intelligent tree growth strategy might be slightly slower to develop meaning structure than the probabilistic tree growth strategy.
- the intelligent tree growth strategy is based on a *simple heuristic* (Gigerenzer & Todd, 1999), in that it does not consider every possible λ in every possible future discrimination game, and then decide which would be the optimal node to refine, but searches only until it finds one which satisfies the chosen criterion.

For reference, I summarise below the definitions of the two tree growth strategies which are used in the experiments:

Tree Growth Strategies

Probabilistic Tree Growth, in which each channel has an innate bias which represents the probability of choosing the channel. The agent chooses a random number r, and finds the channel within whose bounds r falls, as shown in figure 8.1.

Intelligent Tree Growth, in which the agent again has channels with innate biases, but they do not represent the probability of being selected, and are used merely to order the channels before searching occurs. Having ordered them, the agent searches through the channels, testing possible leaf nodes until it finds a node $\lambda(c)$ on channel c, which, if it had been already refined, would have led to the current discrimination game being a success.

First, then, let us consider results from experiments which explore the levels of meaning similarity and communicative success under various configurations of the different cognitive biases and tree growth strategies I described above, with particular focus on the following questions:

- do the cognitive biases and tree growth strategies manifest themselves as differences in the way in which agents create meanings and communicate with each other?
- with which combination of tree growth strategy and bias allocation do agents produce the highest (and lowest) levels of meaning similarity and of communicative success?
- does communicative success rely on the need for agents to *share* the same biases, as implicitly suggested by the proposed universality (among humans) of cognitive biases like the *shape bias*?

In all experiments, the feature values of the objects in the world are initialised at the start of the experiment, and are thereafter permanently fixed; in this first set of experiments, representing the basic model to which later ones will be compared, the feature values are taken from a uniform random distribution.

8.2.3 Probabilistic Tree Growth

Probabilistic Tree Growth based on Uniform Biases

In figure 8.2, I show how the level of meaning similarity σ progresses during 1000 discrimination games in which the agents are building their conceptual structure, with different numbers of sensory channels available to them; these results are further summarised in table 8.2. As we would expect from our knowledge of the meaning creation process, most of the workload of meaning creation occurs in the initial part of the simulation,

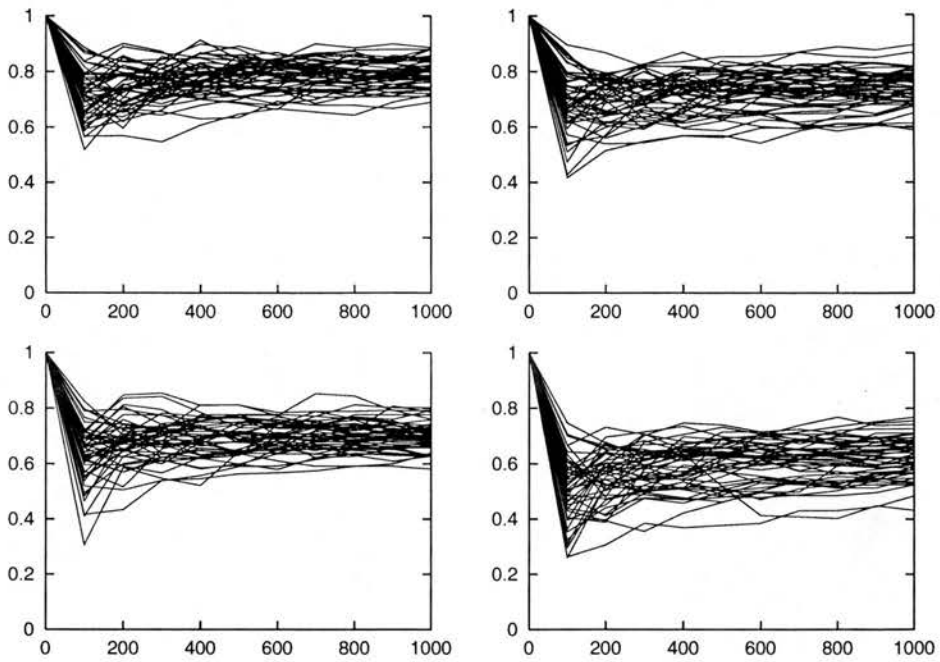


Figure 8.2: Meaning similarity σ in a random world: agents have different experiences and create individual meaning structures using the probabilistic tree growth strategy based on *uniform* channel biases. The simulation contains 1000 discrimination games, and is repeated 50 times, with each run represented by a separate overlaid line on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right)

which leads to dramatic falls[4] in the level of meaning similarity σ at the beginning of the experiment, as the agents fail in the discrimination task, then a slower increase over time, before σ remains more or less constant from only a few hundred episodes onwards, as the agent's conceptual structures are developed sufficiently to represent the world they inhabit, and there is no need to create any further meanings.

We can see that the precise parameters of the lines, from the steepness and depth of the intial fall, through the time taken to recover to a stable level, to the value of this level itself, are dependent, as we would expect, on the number of sensory channels available to the agents. There are, moreover, slight variations in the clustering effects, which we can see both in figure 8.2 and table 8.2; the value of the mean $\bar{\sigma}$ decreases as the number of sensory channels grows, and the variation $\mathrm{CoV}(\sigma)$ increases at the same time. Unsurprisingly, given our understanding of the random processes at work, there is a vanishingly

[4]It is worth noting that the level of both tree similarity τ and agent meaning similarity σ are always artificially high ($\tau = 1.0; \sigma = 1.0$) at the beginning of every experiment, because all the agents' trees are without any growth, and trees in this condition are always necessarily identical.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)
2	0.80	(0.78 – 0.81)	0.89	0.69	0.06
3	0.74	(0.72 – 0.76)	0.89	0.59	0.08
5	0.70	(0.69 – 0.71)	0.80	0.58	0.07
10	0.62	(0.61 – 0.64)	0.77	0.43	0.11

Table 8.2: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *uniform* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment.

small chance of agents developing independent, synchronised ($\sigma = 1$) conceptual structures, and levels above 80% are very rarely seen, and only at all when the number of sensory channels is small. If we compare the results with five sensory channels to those we looked at in table 7.4, when the agents were provided with random conceptual structure, we can see that there are small, but interesting differences:

- the average level of meaning similarity $\bar{\sigma}$ is slightly higher at 0.70 (0.69–0.71) compared to 0.66 (0.65–0.67);
- the average level of communicative success $\bar{\kappa}$ is very slightly lower at 0.90 (0.89–0.91) compared to 0.93 (0.92–0.94).

These results initially might appear to be slightly counter-intuitive, as both sets of experiments appear to be set up with random parameters; the only difference between them, indeed, is that in the original experiment, the meanings are created through the random innate provision described in section 7.2, while in these experiments, they are created randomly following the failure of discrimination games[5]. This apparent near-identity between the two experiments, however, hides one important difference in the degree of randomness involved in the meaning creation procedure.

I have already discussed in detail the process of tree creation during random innate provision, and shown mathematically why the levels of meaning similarity are in fact so predictably consistent, so I will confine myself here to a brief investigation of the agent-driven meaning creation process. As we have seen, in both cases, a sensory channel is

[5]I have, of course, been describing the meaning creation process in this section as probabilistic meaning creation based on uniformly distributed channel biases. Choosing a sensory channel under these circumstances is, however, exactly equivalent to 'choosing a channel at random' as we did in the previous chapter. For expository purposes, however, it is helpful to group this process with the other processes which also use probabilistic meaning creation, but which are based on non-uniform distributions of the channel biases.

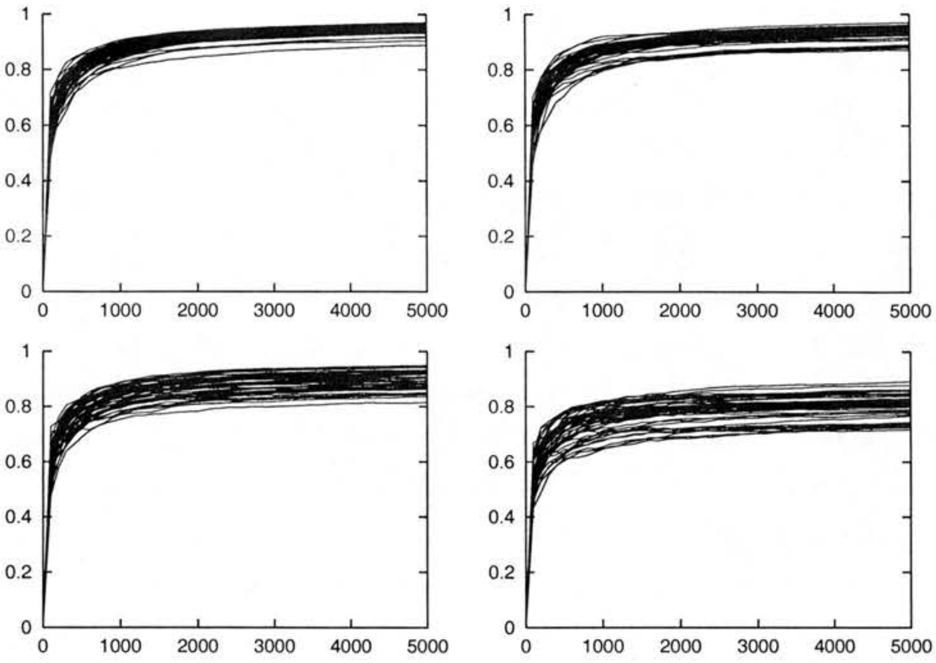


Figure 8.3: Communicative success κ in a random world, following individual meaning creation based on agents' different experiences using the probabilistic tree growth strategy based on *uniform* biases. The simulation contains 5000 communicative episodes, and is repeated 50 times, with each run represented by a separate overlaid line on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right)

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)
2	0.95	(0.94 – 0.95)	0.97	0.89	0.02
3	0.93	(0.93 – 0.94)	0.97	0.87	0.02
5	0.90	(0.89 – 0.91)	0.95	0.82	0.04
10	0.81	(0.80 – 0.82)	0.89	0.72	0.05

Table 8.3: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the probabilistic tree growth strategy based on *uniform* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment.

chosen randomly following the failure of a discrimination game. The difference, however, becomes apparent when we consider how to choose the particular leaf node λ which will be refined. Under innate concept provision, a random value between the bounds of the sensory channel is chosen, and the leaf node corresponding to this value is refined. When the agents create their own meanings, however, the value chosen is taken from the value of the target object itself, as I explained in section 8.2.1; this is in fact how I ensure that the agent's meaning creation process is *grounded* in its experiences of the world. This means that, rather than there being an infinite number of possible values at every step of the meaning creation process, there are in fact only o values which can be chosen at any one time, where o is the number of objects in the world. The probability of a particular node being chosen is not straightforwardly inversely proportional to the depth $d(\lambda)$ of the leaf node λ, or defined by $\frac{1}{2^d}$, as we saw in section 7.4, but is dependent instead on the actual distribution of objects in the world *as well as* the node's depth in the discrimination tree. Because the world, once created, is static, it is inevitable that particular portions of the meaning space, to which no objects correspond, will never be chosen for refinement. The main consequence of this explicit grounding of the new meanings in the agents' environment is, of course, that they reflect the (random) regularities which are found there, and so values are not being chosen from a uniformly random meaning space. The meaning structures created, therefore, will have a slightly greater degree of meaning similarity than under the random innate meaning provision described in chapter 7, but this increase will not be enormous, because the meaning creation is still taking place under the very great pressure for the production of trees which are as balanced and comprehensive as possible (see section 7.4 for more details).

After having created their individual conceptual structures, the agents communicate with each other during 5000 communicative episodes, which are shown in figure 8.3, and summarised in table 8.3. Communicative success κ rises quickly at first, as many lexical items are successfully acquired and understood, but then continues more slowly as the remaining words are only slowly disambiguated from context, in a similar way to that which we saw in the preliminary experiments in section 7.3. We see again that communicative success is very high when there are relatively few lexical items to learn the meanings of, but as the number of sensory channels increases, the average success rate $\bar{\kappa}$ declines and the variation CoV(κ) slowly increases. The scatter plot of meaning similarity σ and communication success κ in figure 8.4 shows both of these phenomena; the crosses appear both lower and less clustered together in the sub-figure at the bottom right.

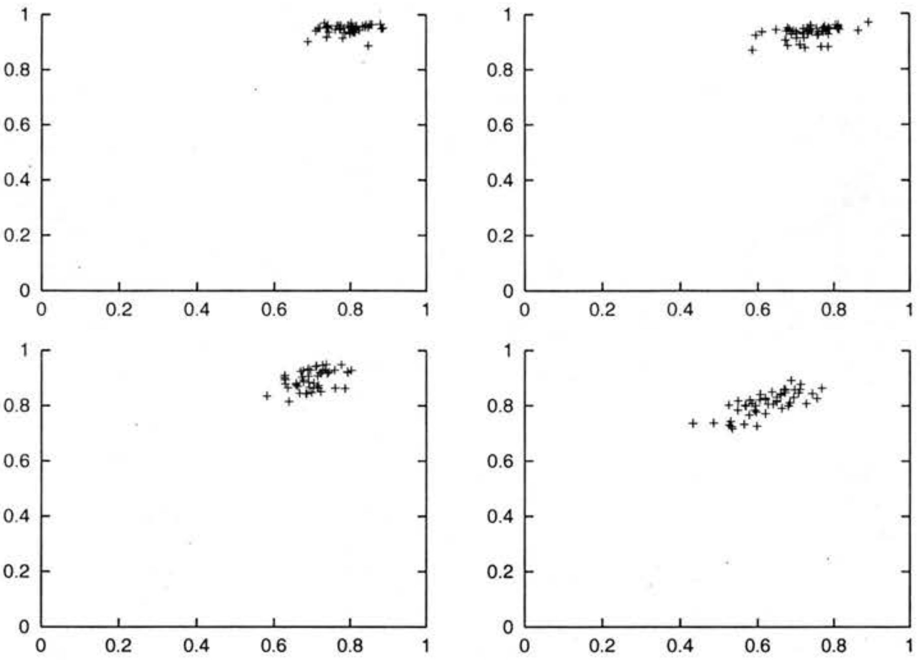


Figure 8.4: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a random world, after 1000 discrimination games and 5000 communicative episodes; the agents have different experiences of the world and create individual meaning structures using the probabilistic tree growth strategy based on *uniform* channel biases. The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

Probabilistic Tree Growth based on Random Biases

Figure 8.5 shows the levels of meaning similarity σ for a similar set of experiments in which the agents create meanings using the probabilistic tree growth strategy based on random biases, and again these experiments are summarised in table 8.4. If we compare these with the results obtained when the agents' probabilistic strategy was based on uniform biases, we can see some marked differences:

- the average $\bar{\sigma}$ is in all cases much lower, and moreover varies very little with respect to the number of channels available;
- there is much more variation in the individual levels of σ achieved on a particular run, shown both by the wide spread of lines in figure 8.5 and by the high values of CoV(σ) in table 8.4; this variation is most pronounced with small numbers of sensory channels available;
- the distributions are very significantly different from those seen in figure 8.2.

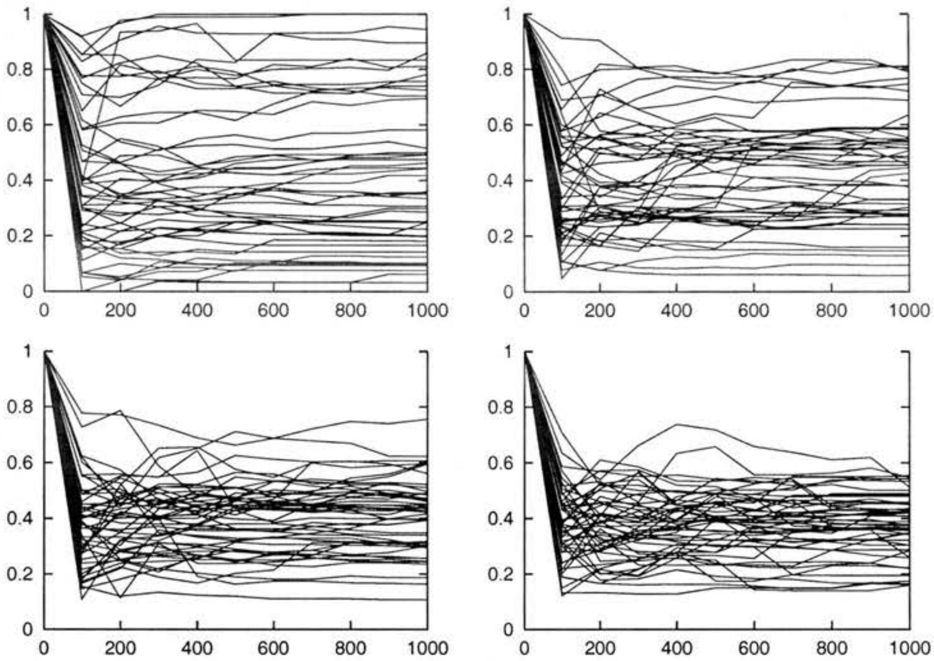


Figure 8.5: Meaning similarity σ in a random world: agents have different experiences and create individual meaning structures using the probabilistic tree growth strategy based on *random* channel biases. The simulation contains 1000 discrimination games, and is repeated 50 times, with each run represented by a separate overlaid line on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right)

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.2)
2	0.42	(0.34 – 0.50)	1.00	0.00	0.66	0.76 **
3	0.44	(0.38 – 0.49)	0.81	0.06	0.44	0.82 **
5	0.41	(0.37 – 0.44)	0.76	0.11	0.32	0.96 **
10	0.38	(0.35 – 0.41)	0.55	0.16	0.29	0.86 **

Table 8.4: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *random* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.2, as shown in the far right-hand column.

The reduced levels of σ with tree growth based on random biases can be explained if we concentrate on the different levels at which randomness operates in the world. We saw in section 7.4 that, at the tree level, random refinements provide a universal pressure to create relatively comprehensively refined trees. Given two trees refined the same number of times, we can predict an expected level of similarity τ between them. But at the *agent* level, there is no such pressure for predictability at all; the randomly distributed cognitive biases will by their nature be completely different for different agents, and so it is not unlikely at all that one agent will have a very low bias for, say, *smell*, but that another agent will have a very high bias for the same channel. Remember that meaning similarity at the agent level σ is calculated by comparing trees on the same sensory channel with each other, as detailed in equation 5.15. High levels of σ are only possible if the vast majority of sensory channels themselves have high levels of τ, and under the probabilistic tree growth strategy, levels of τ depend explicitly on the agent's cognitive biases, which define the distribution of refinements amongst the sensory channels. We could suggest, therefore, that the agents are only likely to produce high levels of meaning similarity σ under the probabilistic tree growth strategy, if their innate biases are similar in the first instance.

Meaning similarity is substantially reduced if we compare random biases with uniform biases under the probabilistic strategy, but what happens to the level of communicative success κ? In figure 8.6, we can see the progression of κ develops in these same experiments, and see summary statistics in table 8.5. The average communicative success $\bar{\kappa}$ is also reduced, with a corresponding increase in the variation CoV(κ). Both the average and the variation in communicative success are reasonably constant across the number of sensory channels, although as we would expect, κ is slightly lower as the agents have more semantic hypotheses to disambiguate. The level of communicative success is substantially higher than the level of meaning similarity in almost all cases, as can be seen in figure 8.7, where the points on the scatter plot are overwhelmingly, with only very few exceptions, above the main diagonal where $\sigma = \kappa$. We saw in chapter 7.5 that this is due to the Gricean maxims of communication which promotes the use of objectively more general meanings in the communicative process. As more general meanings are more likely to be shared by the agents, they are more likely to result in successful communicative episodes, and if they are disproportionately used, then the average level of communicative success $\bar{\kappa}$ will always be higher than the average level of meaning similarity. In figure 8.7, the relationship between σ and κ is perhaps more obvious than it was in figure 8.4 because of the much wider variation in the values of σ, but we have no reason to suppose that it does not hold in all these experiments.

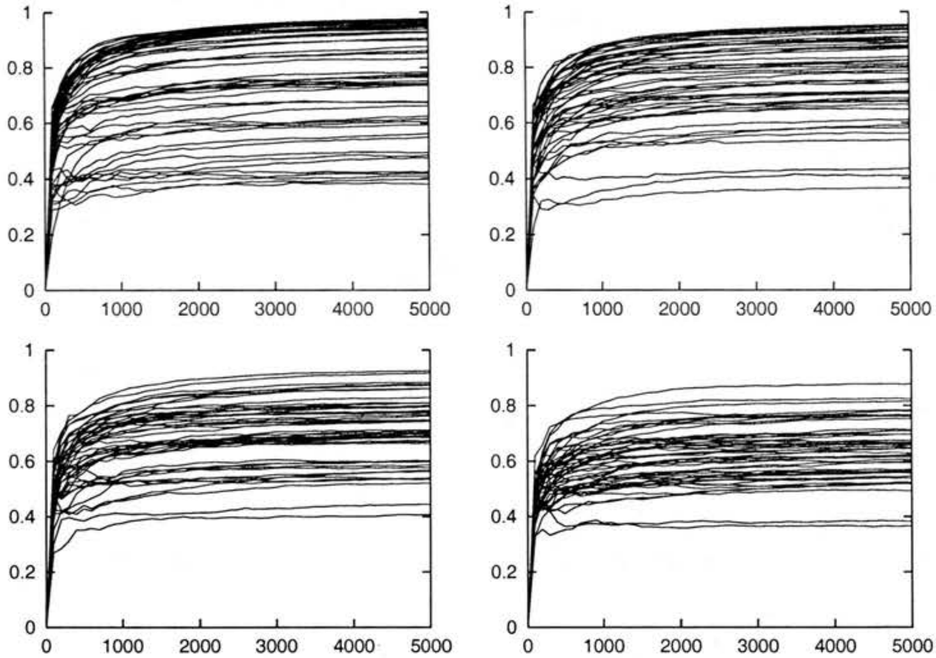


Figure 8.6: Communicative success κ in a random world, following individual meaning creation based on agents' different experiences using the probabilistic tree growth strategy based on *random* biases. The simulation contains 5000 communicative episodes, and is repeated 50 times, with each run represented by a separate overlaid line on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right)

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.3)
2	0.76	(0.71 – 0.82)	0.98	0.38	0.25	0.62 **
3	0.77	(0.73 – 0.81)	0.95	0.37	0.19	0.70 **
5	0.71	(0.68 – 0.74)	0.93	0.41	0.17	0.84 **
10	0.64	(0.61 – 0.67)	0.88	0.37	0.16	0.80 **

Table 8.5: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the probabilistic tree growth strategy based on *random* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.3, as shown in the far right-hand column.

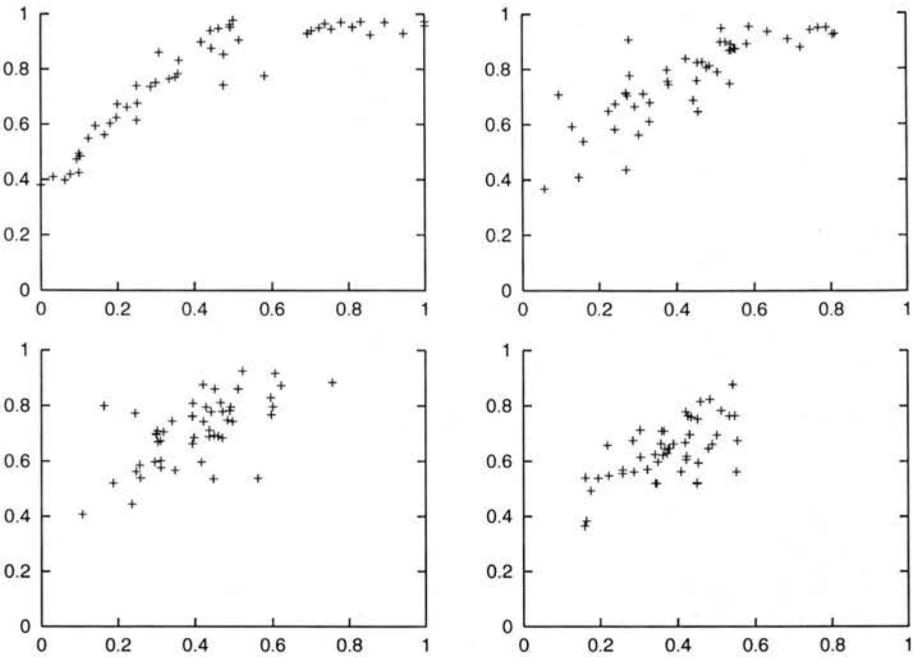


Figure 8.7: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a random world, after 1000 discrimination games and 5000 communicative episodes; the agents have different experiences of the world and create individual meaning structures using the probabilistic tree growth strategy based on *random* channel biases. The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

Probabilistic Tree Growth based on Proportional Biases

The random assignment of biases, therefore, appears to be a confounding factor which undermines, to a large extent, any potential advantage which cognitive biases in themselves may provide to the agents, because the agents are likely to have *different* cognitive biases, which leads to meaning structures which are also too different to allow very successful communication. To investigate what happens when we explicitly set the biases using the proportional method described in section 8.2.1, we look at table 8.6, which summarises the meaning similarity rates in a similar set of experiments, but for agents whose cognitive biases are set proportionally, with $(p = 0.5)^6$.

The results show a swing back to those we found in figure 8.2, when the cognitive biases were set in a uniform distribution; the results here are indeed very similar to those in table 8.2, with the exception that when a large number of channels is available, the variation

[6] For a graphical view of these results, please refer to appendix C, which contains comprehensive details on all the experiments in this chapter.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.2)	KS(8.4)
2	0.80	(0.78 – 0.82)	0.92	0.60	0.07	0.12	0.74 **
3	0.76	(0.74 – 0.77)	0.84	0.63	0.07	0.24	0.84 **
5	0.71	(0.69 – 0.73)	0.82	0.50	0.08	0.20	0.94 **
10	0.71	(0.68 – 0.74)	0.91	0.47	0.14	0.50 **	0.92 **

Table 8.6: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *proportional* ($p = 0.5$) channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.2 and 8.4, as shown in the far right-hand columns.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.3)	KS(8.5)
2	0.95	(0.95 – 0.96)	0.97	0.89	0.01	0.12	0.64 **
3	0.93	(0.93 – 0.94)	0.97	0.89	0.02	0.14	0.78 **
5	0.90	(0.89 – 0.91)	0.96	0.81	0.04	0.16	0.82 **
10	0.85	(0.83 – 0.87)	0.95	0.72	0.06	0.42 **	0.82 **

Table 8.7: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the probabilistic tree growth strategy based on *proportional* ($p = 0.5$) biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.3 and 8.5, as shown in the far right-hand columns.

is greater, and the distributions are significantly different. In comparison with table 8.4, on the other hand, it is clear, unsurprisingly, that using proportional biases reduces the level of variation and increases meaning similarity. In table 8.7, we can see that communicative success, too, is improved dramatically in comparison with randomly set biases, and is again significantly different from the results obtained with uniform biases when ten channels are available.

It is clear from these initial investigations that we can differentiate the uniform and proportional bias allocations, on the one hand, from the random bias allocation on the other; Under uniform and proportional bias allocation, σ varies between 60% and 80% depending on the number of sensory channels available, and κ varies likewise between 80% and 95%, but under random bias allocation, both levels are significantly lower, σ around 40%, and κ between 65% and 75%, so using random biases under the probabilistic tree growth strategy puts the agents at a disadvantage in terms of increasing co-ordination of semantic structure. The most obvious difference between these two groups of bias allocations

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.2)	KS(8.4)	KS(8.6)
2	0.81	(0.76 – 0.85)	1.00	0.50	0.21	0.36 **	0.68 **	0.34 **
3	0.71	(0.68 – 0.75)	1.00	0.45	0.18	0.26	0.66 **	0.32 **
5	0.63	(0.60 – 0.67)	0.91	0.31	0.19	0.44 **	0.70 **	0.48 **
10	0.61	(0.58 – 0.64)	0.78	0.31	0.17	0.20	0.76 **	0.44 **

Table 8.8: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *identical random* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.2, 8.4 and 8.6, as shown in the far right-hand columns.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.3)	KS(8.5)	KS(8.7)
2	0.95	(0.94 – 0.96)	0.98	0.83	0.04	0.48 **	0.56 **	0.44 **
3	0.92	(0.90 – 0.93)	0.98	0.75	0.05	0.30 *	0.58 **	0.34 **
5	0.89	(0.87 – 0.90)	0.97	0.72	0.06	0.14	0.76 **	0.14
10	0.85	(0.83 – 0.86)	0.94	0.73	0.06	0.38 **	0.84 **	0.12

Table 8.9: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the probabilistic tree growth strategy based on *identical random* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.3, 8.5 and 8.7, as shown in the far right-hand columns.

is that in the first group (uniform and proportional), the agents are guaranteed to have the *same* cognitive biases, whereas in the second group they are vanishingly unlikely to have the same biases. In the following section, therefore, I modify the random bias allocation method to test this hypothesis further.

Probabilistic Tree Growth based on Identical Random Biases

In order to test the validity of the hypothesis that the crucial difference between uniform and proportional biases on the one hand, and random biases on the other is the identity of the cognitive biases across agents, I alter the allocation of random biases to ensure that both agents have the same set of cognitive biases; the random biases are created in the same way as before, but each set of biases is created only once, and given to both agents, rather than each agent's biases being generated separately. If the hypothesis is correct, we should expect that agents with identical random biases can also communicate with a similar level of success as those whose biases have been allocated uniformly or proportionally. In table 8.8, however, we find an intriguing set of results. As expected, we can indeed see that the levels of meaning similarity σ are, in all cases, significantly higher than with different random biases (table 8.4). On the other hand, although they are certainly closer, the levels of σ with identical random biases are also still lower in most cases than those with proportional (table 8.6) and uniform (table 8.2) biases. This can be explained if we remember that the actual setup of cognitive biases is actually very different in each of these cases; with identical random biases, each run of the experiment is set up differently, and we therefore find a much higher level of variation in σ. The proportional and uniform allocations, on the other hand, are deterministic, and so each run under these conditions is set up identically, leading to much less potential (and actual) variation. This difference in variation CoV(σ), rather than the average value $\bar{\sigma}$, shows up in the KS test as a significantly different distribution of results.

We will not be surprised, given the consistent relationship between meaning similarity and communicative success which we have seen throughout these experiments, to find a similar story in table 8.9 with the level of communicative success κ. Again, the most significant differences are between identical random biases and different random biases (table 8.5), but there are also significant differences in many of the other conditions as well, particularly when few channels are available.

Summary

It is clear, therefore, that under the probabilistic tree strategy as a whole, we can draw the following conclusions:

- variation in σ is very small if the agents have uniform or proportional biases, but is much wider if the agents have random biases;
- variation in κ is always much smaller, as agents can learn to communicate even with different conceptual structures;
- there is a strong, J-shaped[7] relationship between the levels of σ and κ in all cases, although it is more obviously visualised in some experiments where wide variations in σ are found (see for example figure 8.7) than in experiments with little variation in σ (see for example figure 8.4);
- most importantly, if all agents have the *same* cognitive biases, they will, on average, produce more similarity in their individually created conceptual structure; high levels of meaning similarity, in turn, lead on average to high levels of communication success.

[7]The curve resembles a letter J reflected in the $x = y$ axis: as x (σ) increases from 0 to 1, y (κ) increases more rapidly at first, then slows down substantially.

It is important to note that having the same cognitive biases does not guarantee that agents will build the same conceptual structure, but I have shown through the experiments in this section that they will build conceptual structures which are, on average, more similar. Likewise, agents with high levels of meaning similarity are much more likely to be able to communicate with high levels of success; under the probabilistic tree growth strategy, we can now be confident that this is much more likely to happen if they start off with the same cognitive biases.

In this section, therefore, we have seen that under the probabilistic tree growth strategy, the highest levels of both meaning similarity σ and communicative success κ occur when the agents have identical biases, though the actual expected levels are dependent on the particular bias allocation mechanism and the number of sensory channels available to the agents. Remember that these results are not intended to suggest, in themselves, that *particular* cognitive biases are better than other cognitive biases, so we cannot say whether, for instance, Landau et al. (1988)'s *shape bias* explains more of the lexicon acquisition problem than Markman and Hutchinson (1984)'s *taxonomic bias*, but they do strongly suggest that the *sharing* of cognitive biases, and therefore their universality, whatever their actual realisation, is crucial in substantially increasing the agents' meaning similarity. This is in turn instrumental in higher communicative success rates, and therefore such sharing of cognitive architecture could be a very important pre-adaptation for the emergence of communication.

8.2.4 Intelligent Tree Growth

The intelligent tree growth strategy differs from the probabilistic strategy in that it is not determined by the agent's biases, as I discussed in section 8.2.2. The agents do make use of their cognitive biases to a small extent, in that they order the channels according to the biases before searching through to find a suitable candidate for refinement, but the crucial work of the strategy, and the reason I call it *intelligent*, is that the agent checks the leaf node λ which categorises the target object from each channel in turn, until it finds one which, if it had been refined a further level, would have been successful in discriminating the target object from the other objects in this particular discrimination game. If the agent finds such a leaf node, then it is refined; if none is found, then no refinement at all takes place. The intelligent tree growth strategy therefore almost completely eliminates the random element in the channel selection task, and links meaning creation even more closely to the discrimination process. Not only is tree growth driven by discrimination failure (as of course are all the strategies), but the chosen channel is more likely to be of some use in the future.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.2)
2	0.86	(0.82 – 0.91)	1.00	0.43	0.17	0.64 **
3	0.58	(0.53 – 0.64)	0.95	0.04	0.35	0.58 **
5	0.46	(0.41 – 0.51)	0.89	0.12	0.42	0.76 **
10	0.37	(0.33 – 0.40)	0.72	0.11	0.32	0.88 **

Table 8.10: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the *intelligent* tree growth strategy based on *uniform* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.2, as shown in the far right-hand column.

Intelligent Tree Growth based on Uniform Biases

Let us first take a look at the intelligent tree growth strategy in action, when it is used in conjunction with uniform biases. If we compare the levels of meaning similarity shown in table 8.10 to the corresponding information under the probabilistic strategy (table 8.2), we can see some interesting and significant results. The average meaning similarity $\bar{\sigma}$ is reduced quite substantially, and the variation increased, except when only two channels are available; in this case, $\bar{\sigma}$ is significantly *higher*, and a large number of runs actually produce completely synchronised meaning structures. The reason for this is the nature of the intelligent tree growth strategy, and in particular that refinements are focused on channels which would have succeeded. Other things being equal, channels which already have a high degree of tree growth, and thus a number of specialised meanings, are more likely to produce a discriminatory meaning than those which have only very general meanings. Therefore, after a few initial refinements have been made, the intelligent strategy tends to concentrate refinements on those channels on which trees have already been grown, and moreover on those parts of trees which are refined deeply. Divergence in the conceptual structure is therefore almost inevitable unless the initial refinements made by the agents happen to be the same. If the initial refinements *do* happen to be the same, the intelligent strategy is likely to keep the agents' conceptual structures very *similar* for the same reason: because refinements are concentrated on those areas of the trees on which growth is the deepest, namely those on which the (same) initial refinements were made. The intelligent strategy, therefore, can maintain extreme levels of σ for quite some time. In particular, if there are only two or three sensory channels available, then the initial tree growth is reasonably likely to occur on the same channel in each agent, and so high levels of meaning similarity can ensue. Conversely, if there are many channels available, then

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.3)
2	0.94	(0.92 – 0.96)	0.98	0.58	0.07	0.44 **
3	0.80	(0.76 – 0.84)	0.97	0.35	0.17	0.56 **
5	0.72	(0.68 – 0.76)	0.94	0.42	0.19	0.66 **
10	0.60	(0.57 – 0.62)	0.74	0.42	0.12	0.94 **

Table 8.11: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the *intelligent* tree growth strategy based on *uniform* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.3, as shown in the far right-hand column.

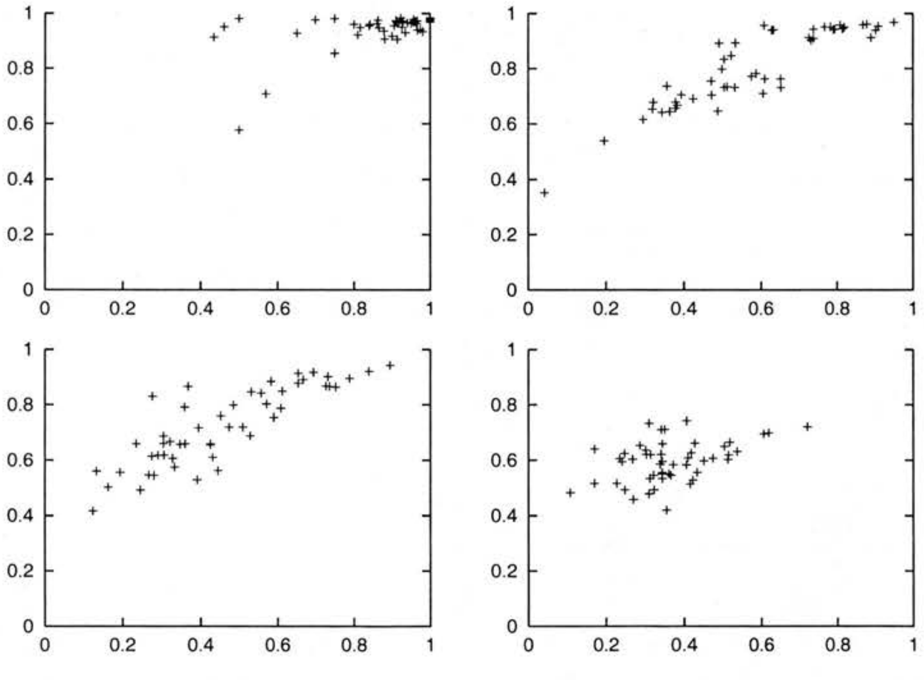


Figure 8.8: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a random world, after 1000 discrimination games and 5000 communicative episodes; the agents have different experiences of the world and create individual meaning structures using the *intelligent* tree growth strategy based on *uniform* channel biases. The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Subfigures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

it is extremely unlikely that such growth will be mirrored across both agents, and so we find very low levels of meaning similarity.

Given these findings on the levels of meaning similarity, it is perhaps not surprising that we find in table 8.11 that the levels of communicative success κ are also significantly lower than those we found in the relevant corresponding experiments where the agents were using probabilistic tree growth (table 8.3). Communication episodes still succeed more often than not, because of the agents' Gricean tendency to prefer general meanings which can be more easily disambiguated through context, but whenever specialised meanings *are* used, it is likely that the hearer will *not* have the speaker's meaning in its conceptual structure, and so may well misidentify the intended referent and fail to understand the utterance. As we have seen, specialised meanings are not only more likely to occur in the conceptual structure under the intelligent tree growth strategy, but moreover are *necessarily* useful in some circumstances, and so are relatively more likely to be used by speakers than specialised meanings which happen to have been created under the probabilistic tree growth strategy, but may well never be called upon. However, the relationship between meaning similarity and communicative success as shown in figure 8.8 is as strong as ever, and possibly even more marked than under the probabilistic strategy, particularly in the experiments with a wide spread of values for σ (see, for instance, the results with 3 and 5 sensory channels); we can see clearly that good communication is very dependent on having high levels of meaning similarity.

Intelligent Tree Growth based on Random Biases

Under the probabilistic tree growth strategy, we saw a large drop in the level of meaning similarity when the agents' biases were allocated randomly, but when such experiments are run under the intelligent tree growth strategy, there is no such drop; instead, we get almost identical results with uniformly and randomly allocated biases, both in terms of meaning similarity (table 8.12 in comparison to table 8.10) and communicative success (table 8.13 in comparison to table 8.11). The average level of meaning similarity $\bar{\sigma}$ is very high when only two channels are available, and is significantly higher than under the probabilistic strategy with random biases (table 8.4), but this decreases rapidly once again, with high levels of variation, when more channels are available. The initial refinements are the most important under the intelligent tree growth strategy, and all refinements are useful to some degree, so it is no real surprise that when there are few channels, there is little scope for variation in the refinements which can take place, and so consequently a high level of meaning similarity.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.4)	KS(8.10)
2	0.82	(0.76 – 0.87)	1.00	0.36	0.23	0.58 **	0.14
3	0.60	(0.55 – 0.66)	0.99	0.22	0.34	0.42 **	0.14
5	0.43	(0.39 – 0.48)	0.70	0.08	0.35	0.22	0.18
10	0.43	(0.39 – 0.47)	0.68	0.17	0.29	0.29 *	0.32 *

Table 8.12: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the *intelligent* tree growth strategy based on *random* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.4 and 8.10, as shown in the far right-hand columns.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.5)	KS(8.11)
2	0.92	(0.89 – 0.94)	0.99	0.57	0.11	0.44 **	0.18
3	0.81	(0.77 – 0.85)	0.97	0.43	0.18	0.22	0.14
5	0.71	(0.68 – 0.75)	0.93	0.36	0.18	0.12	0.12
10	0.61	(0.57 – 0.65)	0.89	0.36	0.21	0.29 *	0.29 *

Table 8.13: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the *intelligent* tree growth strategy based on *random* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.5 and 8.11, as shown in the far right-hand columns.

In the same way, the communicative success rate levels in table 8.13 show very little deviance from those we saw in table 8.11, when the agents had uniform biases; the only significant difference, indeed, is caused by the higher variation in the level of κ when ten channels are available (CoV(κ, random) $= 0.21$, CoV(κ, uniform) $= 0.13$). As we might expect, there is the same clear correlation between levels of meaning similarity and communicative success we have found throughout the model.

The most remarkable fact, indeed, about all the results under the intelligent tree growth strategy is that they are very similar, with few significantly different results[8]. The near-identity of all the results under the intelligent tree growth strategy is very interesting;

[8]The only significant difference to be found is that uniform biases result in significantly lower levels of both meaning similarity and communicative success when ten sensory channels are available, in comparison not only with random biases, as we have just seen in tables 8.12 and 8.13, but also in comparison with proportional and identical random biases as well. For detailed information about these and other experiments, the reader is referred instead to figures C.5-C.8 in appendix C, where full details are collated and reproduced for reference.

although we found that the sharing of cognitive biases by agents was important for communicative success under the probabilistic tree growth strategy, we must conclude now that, by contrast, this same sharing is completely *unimportant* under the intelligent tree growth strategy, which, despite its name, produces significantly lower levels of both σ and κ in all the experiments we have so far looked at.

8.2.5 Summary

Tables 8.14 and 8.15 summarise the experiments carried out so far in this chapter, and show clearly that the levels of meaning similarity σ and communicative success κ which are achieved under both tree growth strategies (probabilistic and intelligent) and all four separate bias allocations (uniform, proportional, (different) random and identical random). Due to considerations of space, I have only included details for the simulations involving the standard five sensory channels in tables 8.14 and 8.15, but comprehensive information for both fewer and more sensory channels is provided for reference in appendix C.

We can see clearly that, under the intelligent tree growth strategy, there are only very small and insignificant differences in the levels of meaning similarity and communicative success which are achieved with different cognitive bias allocations; $\bar{\sigma}$ varies only between 0.43 and 0.47, $\bar{\kappa}$ between 0.71 and 0.75. The allocation of particular cognitive biases has no effect because meaning creation using the intelligent strategy is based almost completely on effectiveness in discrimination, and hardly takes any account of the biases at all. Instead, I have shown that the most important factor in terms of meaning similarity under the intelligent tree growth strategy is the tree growth which has already occurred. If we track backwards through time to the initial scenario in which the agents have no meanings on their conceptual apparatus and are effectively *tabulae rasae*, it would seem reasonable to hypothesise that the level of meaning similarity should be affected by the early experiences which kick-start concept creation, and on which all future concept creation is dependent. The investigation of this hypothesis of the importance of the agents' experience is the subject of section 8.3.

From the experiments in section 7.4 and in this chapter, summarised in tables 8.14 and 8.15, we can confidently conclude that the two tree growth strategies have very different, opposing effects. Under both strategies, we have repeatedly seen that there is an important relationship between the relative levels of meaning similarity and communicative success; we should not be surprised by this, as we already understand that the sharing of conceptual structure is crucial to the inference of the meaning of utterances through

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)
Probabilistic	Uniform	0.70	(0.69 – 0.71)	0.80	0.58	0.07
	Proportional	0.71	(0.69 – 0.73)	0.82	0.50	0.08
	Random	0.41	(0.37 – 0.44)	0.76	0.11	0.32
	Identical Random	0.63	(0.60 – 0.67)	0.91	0.31	0.19
Intelligent	Uniform	0.46	(0.41 – 0.51)	0.89	0.12	0.42
	Proportional	0.46	(0.41 – 0.51)	0.83	0.19	0.37
	Random	0.43	(0.39 – 0.48)	0.70	0.08	0.35
	Identical Random	0.47	(0.42 – 0.51)	0.85	0.13	0.35

Table 8.14: Meaning similarity σ — summary for agents in a random world, with different experiences. The table shows a summary of the final range and distribution of σ across 50 repetitions of each experiment, when agents have five sensory channels.

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)
Probabilistic	Uniform	0.90	(0.89 – 0.91)	0.95	0.82	0.04
	Proportional	0.90	(0.89 – 0.91)	0.96	0.81	0.04
	Random	0.71	(0.68 – 0.74)	0.93	0.41	0.17
	Identical Random	0.89	(0.87 – 0.90)	0.97	0.72	0.06
Intelligent	Uniform	0.72	(0.68 – 0.76)	0.94	0.42	0.19
	Proportional	0.74	(0.71 – 0.78)	0.95	0.42	0.17
	Random	0.71	(0.68 – 0.75)	0.93	0.36	0.18
	Identical Random	0.75	(0.72 – 0.79)	0.96	0.53	0.16

Table 8.15: Communicative success κ — summary for agents in a random world, with different experiences. The table shows a summary of the final range and distribution of κ across 50 repetitions of each experiment, when agents have five sensory channels.

multiple contexts. The differences between the strategies appear to be primarily in the concept creation phase which precedes and provides the foundation for communication, as I summarise below:

Probabilistic Tree Growth, on the one hand,

- exerts a pressure towards the creation of balanced, comprehensively refined discrimination trees on the sensory channel; these discrimination trees will therefore contain more general meanings and relatively few specific meanings;
- the sharing of cognitive biases plays an important role under this strategy; agents with identical cognitive biases will always, on average, end up with higher levels of meaning similarity than those with different biases.

Intelligent Tree Growth, on the other hand,

- exerts a pressure in the opposite direction, towards the creation of imbalanced, deeply skewed discrimination trees, which contain many specific meanings and relatively few general meanings;
- the sharing of cognitive biases under this strategy is consequently unimportant, and has no effect on the level of meaning similarity;
- moreover, the level of meaning similarity is substantially lower under the intelligent tree growth strategy than with the same cognitive bias allocation under the probabilistic tree growth strategy.

8.3 Experience

The model of empirical meaning creation which we have been investigating is explicitly based on the agents' building of their conceptual structure in response to failures in their interactions with the world through the discrimination game. I also hypothesised in section 8.2.4 that, under the intelligent tree growth strategy, the low level of meaning similarity σ was likely to be due to the different experiences the agents had.

In human language communities, it is well known that groups of people who have similar experiences create specialised semantic distinctions based on those experiences, leading to the creation of particular lexical terminology or jargon to name the distinctions they have made. The distinctive styles of legal documents or medical terminology are perhaps the most well-known (and most widely mocked), but in fact all groups who share experiences create such distinctions. Despite the pejorative connotations of the word 'jargon', indeed, such specialised terminology is not only essential for making the distinctions which are important to the group, but are also actually very efficient and economical within the context of the group (Allen, 2001), even if they are often seen as obfuscatory outwith that community. In section 3.4, indeed, we saw that the semantic categorisations and classifications made by speakers of different languages are remarkably varied, and concluded that a grounded mechanism of meaning construction was an essential part of this model. In this section, therefore, I investigate the importance of the specific situations which are experienced by the agents, and in particular how much of their conceptual structure is influenced by the order in which they encounter certain target objects in context, by looking at simulations in which both agents play the *same* discrimination games. By comparing the results we obtain under these condition to those we found when the agents had different experiences (in section 8.2), we will be able to come to conclusions about the effects which experience has on both meaning creation and communicative success. In these simulations, a set of objects is chosen as usual for the discrimination game (see section 4.3.1), but this time the agents take it in turn to play the same game individually. In order to maintain our policy of avoiding feedback, neither agent knows that the game has been played before, or that other agents have been exposed to the same experience.

8.3.1 Probabilistic Tree Growth Strategy

If we investigate tables 8.16 and 8.17, which show the rates of meaning similarity achieved under the probabilistic tree growth strategy with uniform and random bias allocation respectively, we see that there are few differences from the results obtained when the agents had different experiences. Although the meaning creation process in all these experiments is grounded to the extent that concept development is triggered by an individual failing to interact with its environment in an appropriate manner, the probabilistic tree growth strategy does not take this grounding any further; once the failure in discrimination has taken place, no further use is made of the context in which the target object was observed, and so there are few differences. There is a significant difference under uniform biases, when only two sensory channels are available, but this appears to be an artefact due to a couple of experiments which produced outlying results; this is much more likely to happen with few channels available because each sensory channel contributes such a high proportion to the overall level of meaning similarity.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.2)
2	0.84	(0.82 – 0.86)	0.96	0.64	0.08	0.50 **
3	0.75	(0.73 – 0.77)	0.92	0.62	0.08	0.14
5	0.70	(0.69 – 0.72)	0.81	0.49	0.09	0.18
10	0.63	(0.60 – 0.65)	0.74	0.43	0.10	0.17

Table 8.16: Meaning similarity σ in a random world, after agents have had 1000 *identical* discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *uniform* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.2, as shown in the far right-hand column.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.4)
2	0.43	(0.36 – 0.50)	1.00	0.00	0.59	0.14
3	0.45	(0.40 – 0.50)	0.92	0.05	0.41	0.12
5	0.45	(0.41 – 0.48)	0.73	0.22	0.29	0.18
10	0.39	(0.36 – 0.42)	0.68	0.22	0.25	0.14

Table 8.17: Meaning similarity σ in a random world, after agents have had 1000 *identical* discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *random* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.4, as shown in the far right-hand column.

Figure 8.9 shows the now familiar relationship between meaning similarity σ and communicative success κ, for agents with random biases under the probabilistic strategy, with the inverted J-shape of the curve being much more obvious when the range of values for σ is more spread out, when few channels are available (e.g. CoV(σ), 2 features $= 0.59$; CoV(σ), 10 features $= 0.25$). Levels of communicative success, too, are broadly similar to the results when the agents had different experiences, but again with some significantly higher and others significantly lower.

Overall, the results for the probabilistic tree growth strategy are at best inconclusive, and we cannot say that giving the agents identical experiences makes any systematic difference to either their conceptual structures or communicative prowess.

8.3.2 Intelligent Tree Growth Strategy

In sharp contrast, table 8.18 shows how meaning similarity levels are significantly increased across the board under the intelligent tree growth strategy with uniform biases,

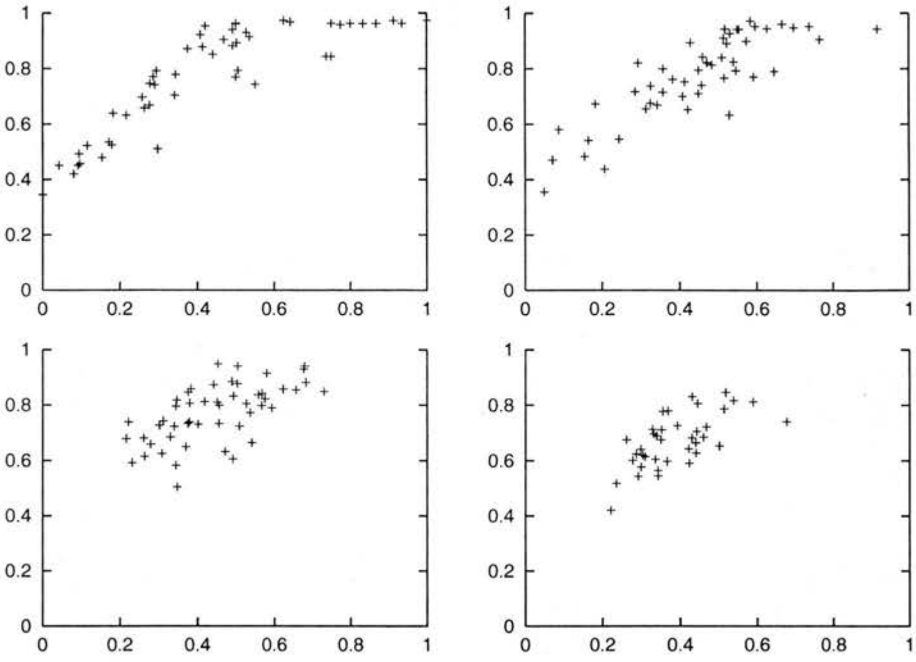


Figure 8.9: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a random world, after 1000 discrimination games and 5000 communicative episodes; the agents have *identical* experiences of the world and create individual meaning structures using the probabilistic tree growth strategy based on *random* channel biases. The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

with almost perfect synchronisation occuring when the agents have few channels. This occurs, of course, because the agents' meaning creation under intelligent tree growth is embedded much more closely into their experience, so when they have the same experiences, they choose the same meanings to refine. But given that the agents have both the same biases and the same experiences, and the latter control the process of meaning creation, why do we only get near-perfect meaning creation, and then only when few channels are available? After all, the agents appear to be in a deterministic situation, but yet do not develop identical conceptual structures. In these experiments, the agents both have exactly the same experiences, and so meaning creation will necessarily be triggered by the failure of the same discrimination game at the start of the simulation; they then both go through their sensory channels, which are ordered identically due to their biases, and both find the same leaf node to refine under the intelligent tree growth strategy.

The answer to this conundrum lies in an oddity which arises because uniform biases are identical not only across agents, but across sensory channels as well. When using the intelligent tree growth strategy, the agent orders its channels according to its cognitive

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.10)
2	0.99	(0.99 – 1.00)	1.00	0.93	0.02	0.74 **
3	0.89	(0.86 – 0.93)	1.00	0.54	0.14	0.66 **
5	0.70	(0.65 – 0.75)	1.00	0.35	0.25	0.48 **
10	0.55	(0.49 – 0.61)	0.94	0.26	0.30	0.55 **

Table 8.18: Meaning similarity σ in a random world, after agents have had 1000 *identical* discrimination games and created individual meaning structures using the *intelligent* tree growth strategy based on *uniform* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.10, as shown in the far right-hand column.

biases, and then searches through them. But if all channels have the same bias, how then can they be ordered? The solution I adopt is to randomly order all channels which have equal biases, which means that the two agents may well search through their identically-biased channels in a different order. It is important to note that this is not absolutely equivalent to choosing a channel at random, because it is not the channel to be refined which is being chosen, but simply the order in which the channels will be searched until a suitable node is found. Although this distinction makes no difference with few channels, leading to almost completely synchronised meaning structures, as the number of channels increases, the random nature of the ordering becomes more prominent, and the level of meaning similarity approaches that which is achieved using different random biases.

The problem, or feature, of uniform biases being equal not just along the dimension of individual agents but also across the dimension of sensory channels within an agent does not, of course, apply either to the proportional or identical random bias allocation strategies. In these circumstances, the determinism of identical biases and identical experiences does indeed lead to perfect meaning sychronisation (with $\sigma = 1$) on every occasion and thence to near-optimal communication success. With different random biases, on the other hand, there is no determinism in the model, and consequently, as we can see in table 8.19, the levels of meaning similarity are much lower. They are, however, still significantly higher than with different experiences, with some simulations producing complete synchronisation even with as many as five sensory channels available.

8.3.3 Summary

In all cases, the high levels of meaning similarity lead inevitably to significantly improved communication when the agents have the same experiences. The experiments in this

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.12)
2	0.99	(0.98 – 1.00)	1.00	0.67	0.05	0.82 **
3	0.87	(0.83 – 0.91)	1.00	0.35	0.17	0.58 **
5	0.61	(0.56 – 0.67)	1.00	0.23	0.33	0.46 **
10	0.51	(0.45 – 0.56)	0.94	0.21	0.29	0.29

Table 8.19: Meaning similarity σ in a random world, after agents have had 1000 *identical* discrimination games and created individual meaning structures using the *intelligent* tree growth strategy based on *random* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.12, as shown in the far right-hand column.

section have shown us that the two different tree growth strategies again have opposing effects on the levels of meaning similarity and communicative success, but that these effects are different when the agents have identical experiences from the effects we saw when experiences were different.

When the agents had different experiences of the world, the probabilistic tree growth strategy's pressure towards balanced trees leads to a reasonable level of meaning similarity, but the intelligent tree growth strategy's concentration on providing specific meanings led to vary diverse meaning structures. By contrast, in the summary tables 8.20 and 8.21, we can see that when the agents have the same experiences of the world, we have seen the results change in different ways:

- identity of experience has no major effect on the results under probabilistic tree growth;
- under intelligent tree growth, however, identity of experience combines with the strategy's pressure to build trees with many specific meanings, so that complete meaning synchronisation ($\bar{\sigma} = 1$) occurs deterministically in every case if the agents have identical biases and identical experiences;
- even if the agents do not have identical biases, meaning similarity rates are significantly increased when the agents have the same experiences of the world, allowing them to learn to communicate successfully much more quickly.

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)
Probabilistic	Uniform	0.70	(0.69 – 0.72)	0.81	0.49	0.09
	Proportional	0.64	(0.60 – 0.67)	0.95	0.38	0.20
	Random	0.45	(0.41 – 0.48)	0.73	0.22	0.29
	Identical Random	0.64	(0.61 – 0.67)	0.90	0.43	0.17
Intelligent	Uniform	0.70	(0.65 – 0.75)	1.00	0.35	0.25
	Proportional	1.00	(1.00 – 1.00)	1.00	1.00	0.00
	Random	0.61	(0.56 – 0.67)	1.00	0.23	0.33
	Identical Random	1.00	(1.00 – 1.00)	1.00	1.00	0.00

Table 8.20: Meaning similarity σ — summary for agents in a random world, with *identical* experiences. The table shows a summary of the final range and distribution of σ across 50 repetitions of each experiment, when agents have five sensory channels.

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)
Probabilistic	Uniform	0.91	(0.91 – 0.92)	0.96	0.83	0.03
	Proportional	0.87	(0.85 – 0.88)	0.97	0.72	0.06
	Random	0.77	(0.74 – 0.80)	0.95	0.50	0.14
	Identical Random	0.89	(0.88 – 0.90)	0.97	0.74	0.05
Intelligent	Uniform	0.86	(0.84 – 0.89)	0.97	0.61	0.10
	Proportional	0.96	(0.96 – 0.96)	0.98	0.93	0.01
	Random	0.79	(0.75 – 0.82)	0.97	0.50	0.16
	Identical Random	0.96	(0.96 – 0.97)	0.98	0.92	0.01

Table 8.21: Communicative success κ — summary for agents in a random world, with *identical* experiences. The table shows a summary of the final range and distribution of κ across 50 repetitions of each experiment, when agents have five sensory channels.

8.4 The Structure of the World

The world in which we live is not uniformly random; indeed, there are many constant properties behind the phenomena we encounter, which can be described in terms of physical and chemical laws. We know, for instance, that unsupported objects will always fall until they reach a lower surface. Scientists can measure the gravitational force which causes this, and moreover, since Newton, we have known that this force is applicable to all bodies, and its magnitude is proportional to the mass of the two bodies under consideration and inversely proportional to the square of the distance between them. Despite knowing these facts, and being able to build from them to Einstein's theory of general relativity and beyond, we also know, in practical terms, that the gravitational field applying to objects in our world does not differ, and is of no use whatsoever in distinguishing objects from each other; in terms of a space of possible worlds containing different levels of gravitational field, all the objects in our world are *clumped* together in one section of the space, where the field is constant.

The structure of the world has been proposed as an explanatory factor for many problems, including the acquisition of lexical vocabulary both in real infants and in simulation models. In section 3.3, I described the proposed *whole-object bias*, and we explored Bloom (2000)'s description of how babies use the regularities in the structure of the environment around them, in particular the properties of objects like *cohesion*, to make sense of the world. In computational simulations built to explain aspects of language evolution, indeed, K. Smith (2003) has shown how compositional communication systems (those where the meaning of a complex signal is made up a function of the meanings of its constituent parts) are more likely to emerge in a population of generalising agents when the environment exhibits a high degree of structure. In this section, I investigate whether the structure of the world they inhabit can have an effect on the meaning similarity and communicative success of the agents. The experiments discussed hitherto have been carried out in a random world, where the objects were created at random and each of their feature values distributed uniformly throughout the meaning space. In these experiments, although I have investigated different kinds of cognitive biases, I have shown no motivation for the existence of the biases, nor for how they might have arisen. If the agents' cognitive biases are more relevant to the world in which the agents live, so that they reflect the structure of that world, what effects will we find?

I introduce, therefore, the notion of a structured or *clumpy* world, where the objects' feature values are clumped together in various ways. In particular, I implement structure in the world by establishing *groups* of objects, where each member of the group has an *identical* feature value for some particular feature[9]. In the randomly-generated world, it is vanishingly unlikely, given the fact that feature values are real numbers with many significant digits, that any two objects will have exactly the same feature value, and so objects are, in the limit, always distinguishable. In the clumpy world, however, the objects in a particular group are defined *a priori* to be indistinguishable from each other on the chosen sensory channel, no matter how many times its discrimination tree is refined, and so the objects can only be distinguished using meanings created on another sensory channel. Compare, for example, the difference between trying to differentiate a number of plain white sheets of A4 paper, and a similar number of students' faces in a lecture theatre; the sheets of paper are analogous to objects in my model which are indistinguishable, but the individual faces are, by contrast, easily recognisable, as the appropriate distinctive categories are created on the 'face' sensory channel.

In the randomly-generated world, in the limit, we could consider each object as a group in itself, with each group containing just one object; in the clumpy world, I define the number of groups arbitrarily according to the number of the sensory channel and number of objects in the world. The number of groups on channel c, $g(c)$ is defined as follows:

$$g(c) = \frac{O}{c+1} \tag{8.2}$$

where O is the number of objects in the world. If there is no exact division, then $g(c)$ is always rounded up to the next whole number, so that we will always produce at least one group on every channel. In a world of 20 objects, therefore, the number of groups on each channel will be as shown in table 8.22. We can see that the channels towards the end of the list have few groups, and so are much less likely to be of any use in a discrimination game, though we also note that none is completely useless, where all objects fall into one group (this would only happen if the agents had more sensory channels than there were objects in the world). The groups are arbitrarily biased so that more distinctions can be made on low-numbered sensory channels, just as the proportionally allocated biases I described in section 8.2 were biased toward low-numbered sensory channels, thus providing the potential for selectionist motivations for the introduction of proportionally allocated biases, though I will not explore these motivations further in this thesis.

[9]The features are of course still abstract, but it is clear that such a mechanism allows for analogies with the gravitational field between objects which I discussed above.

Channel c	0	1	2	3	4	5	6	7	8	9
Groups $g(c)$	20	10	7	5	4	4	3	3	3	2

Table 8.22: Allocation of objects into groups in a clumpy world. The number of groups for a particular sensory channel is defined by equation 8.2.

8.4.1 Probabilistic Tree Growth

If we run the experiments within the confines of a clumpy world, where the agents use the probabilistic tree growth strategy, we find that there are no major differences in the levels of meaning similarity σ achieved, in comparison with those which we found within a randomly-generated world. Tables 8.23 and 8.24, for instance, show results from simulations based on uniform and random biases, and the only significant change occurs when many channels are available and the biases are uniform. We found a similar pattern in section 8.3, when we altered the agents' experience of the world, also to no great effect; we can conclude that environmental factors such as the agents' experience or the structure of the world have no impact on the probabilistic tree growth strategy, which instead leads to similar levels of meaning similarity under all these different circumstances.

If we turn to communication, however, we find a different story entirely, with enormous increases in the communicative success rate κ being returned for all cognitive biases, compared to the same experiments in a randomly-generated world. Table 8.25 shows experimental results for uniform biases, table 8.26 for random biases, and in all cases, the level of communicative success is significantly higher than within a randomly-generated world. Average communicative success rates of over 90% are common if the agents' cognitive biases are the same, with slightly lower rates if the agents' biases differ. For instance, in figure 8.10, we again find very consistently high levels of both meaning similarity and communicative success when the agents have uniform biases, while in figure 8.11, with random biases, we find a much larger spread of values for meaning similarity, as expected, and can see the enormous premium of communicative success over meaning similarity, as all the points on the graphs are considerably higher than the $x = y$ diagonal.

It is clear, therefore, that while setting the experiments in a clumpy world does not improve the level of meaning similarity very much compared to the corresponding experiments in the random world, enormous improvements are seen in communicative success

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.2)
2	0.80	(0.79 – 0.81)	0.88	0.68	0.06	0.14
3	0.75	(0.73 – 0.76)	0.86	0.63	0.08	0.10
5	0.71	(0.70 – 0.73)	0.82	0.56	0.08	0.20
10	0.70	(0.69 – 0.72)	0.79	0.55	0.07	0.48 **

Table 8.23: Meaning similarity σ in a *clumpy* world, after agents have had 1000 different discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *uniform* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.2, as shown in the far right-hand column.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.4)
2	0.45	(0.39 – 0.52)	0.83	0.00	0.52	0.14
3	0.48	(0.42 – 0.54)	0.93	0.04	0.42	0.20
5	0.44	(0.40 – 0.47)	0.81	0.19	0.29	0.14
10	0.44	(0.41 – 0.47)	0.63	0.20	0.22	0.26

Table 8.24: Meaning similarity σ in a *clumpy* world, after agents have had 1000 different discrimination games and created individual meaning structures using the probabilistic tree growth strategy based on *random* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.4, as shown in the far right-hand column.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.3)
2	0.96	(0.96 – 0.96)	0.98	0.91	0.01	0.42 **
3	0.95	(0.95 – 0.96)	0.98	0.89	0.02	0.44 **
5	0.94	(0.94 – 0.95)	0.98	0.88	0.03	0.64 **
10	0.92	(0.91 – 0.93)	0.98	0.80	0.04	0.88 **

Table 8.25: Communicative success κ in a *clumpy* world, after 5000 communicative episodes following meaning creation using the probabilistic tree growth strategy based on *uniform* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.3, as shown in the far right-hand column.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.5)
2	0.85	(0.80 – 0.89)	0.99	0.31	0.20	0.28 *
3	0.84	(0.80 – 0.88)	0.99	0.34	0.18	0.32 **
5	0.82	(0.79 – 0.84)	0.98	0.50	0.12	0.42 **
10	0.80	(0.76 – 0.83)	0.96	0.45	0.14	0.58 **

Table 8.26: Communicative success κ in a *clumpy* world, after 5000 communicative episodes following meaning creation using the probabilistic tree growth strategy based on *random* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.5, as shown in the far right-hand column.

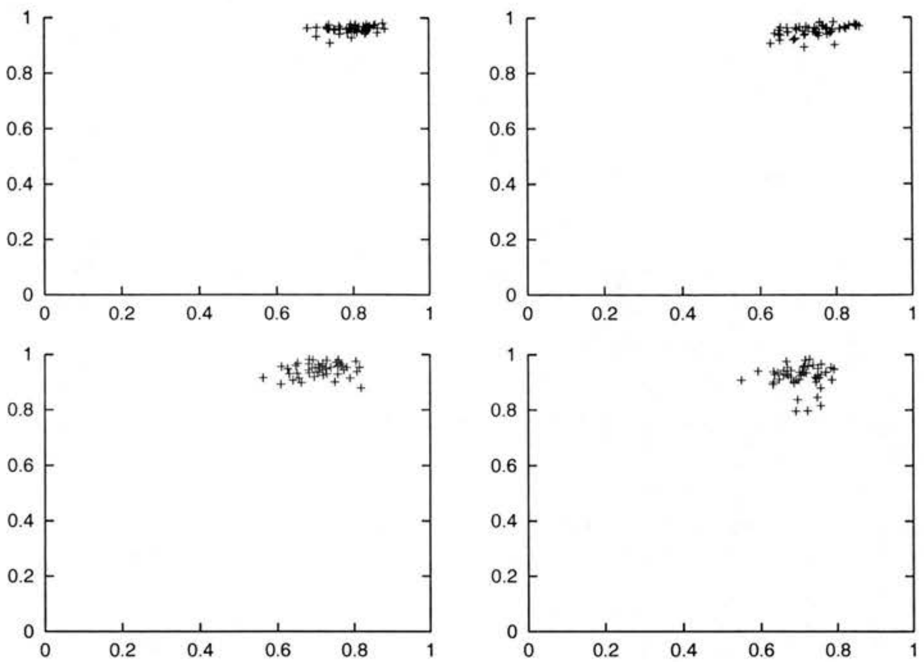


Figure 8.10: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a *clumpy* world, after 1000 discrimination games and 5000 communicative episodes; the agents have different experiences of the world and create individual meaning structures using the probabilistic tree growth strategy based on *uniform* channel biases. The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

levels. In a structured, clumpy world, the hearer's interpretation procedure is made simpler under the introspective obverter methodology because there are generally fewer semantic hypotheses for it to consider; for instance, the hypothetical *gravitational field* feature is very unlikely to produce any possible meanings. Because each particular episode will produce fewer semantic hypotheses to consider, the disambiguation process will take less time and be more successful, which therefore leads to an enormous premium in the level of κ, which is much higher, compared to σ, than in a randomly-generated world.

8.4.2 Intelligent Tree Growth

In the randomly-generated world, we found (see section 8.2) that the intelligent tree growth strategy's pressure to develop meanings on structures which were already considerably refined led to low levels of both meaning similarity σ and communicative success κ. Tables 8.27 and 8.28 show results for uniform and random biases respectively,

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.10)
2	0.94	(0.90 – 0.97)	1.00	0.50	0.13	0.54 **
3	0.93	(0.91 – 0.96)	1.00	0.64	0.11	0.80 **
5	0.83	(0.79 – 0.87)	1.00	0.46	0.17	0.68 **
10	0.81	(0.78 – 0.84)	1.00	0.59	0.12	0.96 **

Table 8.27: Meaning similarity σ in a *clumpy* world, after agents have had 1000 different discrimination games and created individual meaning structures using the intelligent tree growth strategy based on *uniform* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.10, as shown in the far right-hand column.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.12)
2	0.95	(0.92 – 0.98)	1.00	0.50	0.11	0.56 **
3	0.91	(0.88 – 0.95)	1.00	0.56	0.13	0.72 **
5	0.80	(0.77 – 0.84)	1.00	0.53	0.16	0.82 **
10	0.81	(0.78 – 0.84)	1.00	0.56	0.14	0.91 **

Table 8.28: Meaning similarity σ in a *clumpy* world, after agents have had 1000 different discrimination games and created individual meaning structures using the intelligent tree growth strategy based on *random* channel biases. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.12, as shown in the far right-hand column.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.11)
2	0.97	(0.96 – 0.98)	0.99	0.80	0.03	0.38 **
3	0.95	(0.94 – 0.97)	0.99	0.71	0.05	0.72 **
5	0.91	(0.89 – 0.94)	0.98	0.61	0.10	0.70 **
10	0.89	(0.87 – 0.91)	0.98	0.70	0.09	0.92 **

Table 8.29: Communicative success κ in a *clumpy* world, after 5000 communicative episodes following meaning creation using the intelligent tree growth strategy based on *uniform* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.11, as shown in the far right-hand column.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.13)
2	0.96	(0.95 – 0.98)	0.99	0.65	0.06	0.50 **
3	0.94	(0.92 – 0.96)	0.99	0.70	0.07	0.58 **
5	0.91	(0.88 – 0.93)	0.98	0.64	0.08	0.64 **
10	0.89	(0.87 – 0.92)	0.98	0.70	0.09	0.83 **

Table 8.30: Communicative success κ in a *clumpy* world, after 5000 communicative episodes following meaning creation using the intelligent tree growth strategy based on *random* biases. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.13, as shown in the far right-hand column.

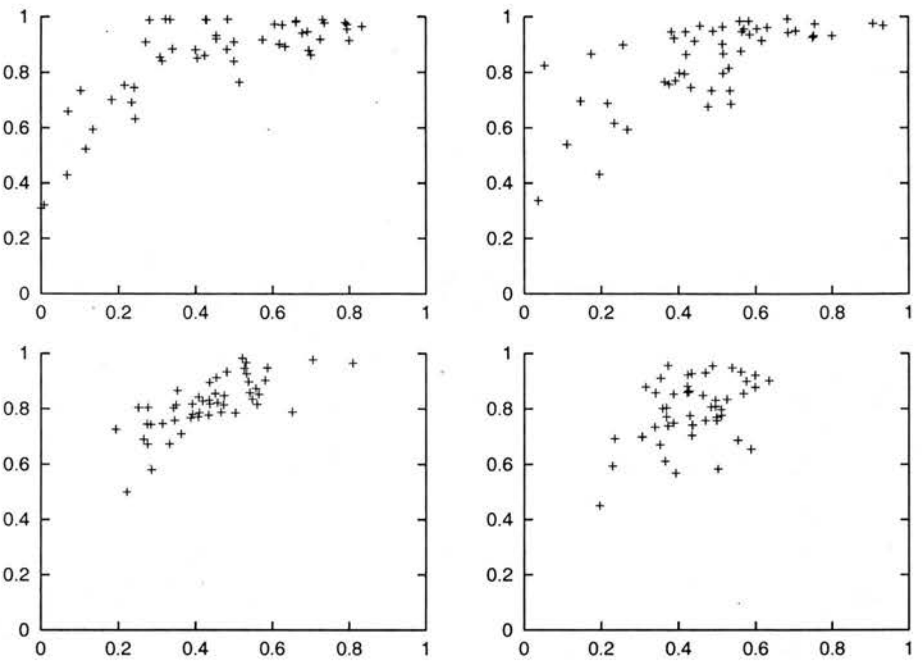


Figure 8.11: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a *clumpy* world, after 1000 discrimination games and 5000 communicative episodes; the agents have different experiences of the world and create individual meaning structures using the probabilistic tree growth strategy based on *random* channel biases. The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

however, set within the context of a clumpy world, and we can see that the levels of meaning similarity are much higher than we might have expected, and massively statistically significantly higher than the results in a random world.

Because the intelligent strategy rejects many possible leaf nodes for refinement because they would not have made a difference in the current discrimination game, agents using this strategy do not waste time and effort growing detailed conceptual structure on sensory channels (or parts of sensory channels) which cannot distinguish between objects in the world. If we consider our hypothetical *gravitational field* feature, for instance, in which all agents have exactly the same feature value, then we can see that the intelligent strategy will always ignore such a sensory channel and will never develop any conceptual structure there. Instead, agents using the intelligent strategy concentrate on those sensory channels which can make a difference; in this way they take account of the structure of the world, in respect of the objects' feature values, or are *ecologically rational*, in Gigerenzer and Todd (1999)'s phraseology. This exploitation of the distribution of the

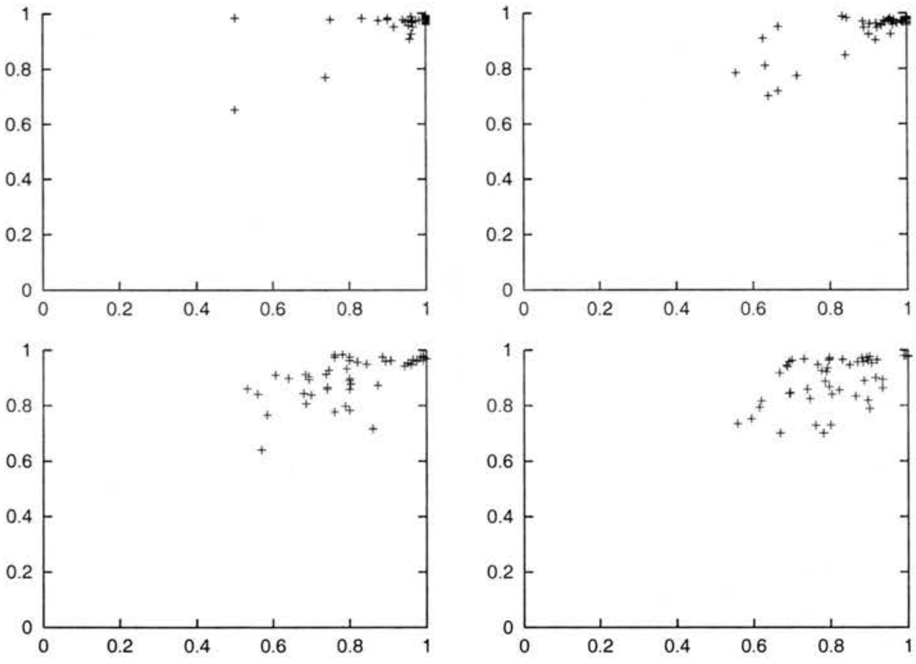


Figure 8.12: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a *clumpy* world, after 1000 discrimination games and 5000 communicative episodes; the agents have different experiences of the world and create individual meaning structures using the *intelligent* tree growth strategy based on *random* channel biases. The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Subfigures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

objects in the world means that the agents do not create unnecessary conceptual distinctions. Because the agents live in the same environment, and therefore both exploit the same environmental structure, over time a much higher level of meaning similarity σ is achieved.

As the levels of meaning similarity are already so much increased when using the intelligent tree growth strategy, the level of communicative success κ cannot possibly show the same enormous premium which we saw under the probabilistic tree growth strategy above. Nevertheless, as we can see in tables 8.29 and 8.30, the levels of communicative success are no less impressive for that, regularly topping 95%, and being significantly higher than in the randomly-generated world in all cases. In figure 8.12, we can see the high levels of both meaning similarity and communicative success, as the points are clumped towards the upper corner of the graph, as well as the familiar relationship between the two variables. Having exploited the information structure in their environment during meaning creation to eliminate unnecessary conceptual growth, the agents now have fewer semantic hypotheses to consider while they are inferring the meaning of an

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)
Probabilistic	Uniform	0.71	(0.70 – 0.73)	0.82	0.56	0.08
	Proportional	0.60	(0.56 – 0.63)	0.95	0.33	0.20
	Random	0.44	(0.40 – 0.47)	0.81	0.19	0.29
	Identical Random	0.68	(0.65 – 0.71)	0.89	0.46	0.16
Intelligent	Uniform	0.83	(0.79 – 0.87)	1.00	0.46	0.17
	Proportional	0.78	(0.73 – 0.83)	1.00	0.33	0.22
	Random	0.80	(0.77 – 0.84)	1.00	0.53	0.16
	Identical Random	0.83	(0.79 – 0.87)	1.00	0.54	0.18

Table 8.31: Meaning similarity σ — summary for agents in a *clumpy* world, with different experiences. The table shows a summary of the final range and distribution of σ across 50 repetitions of each experiment, when agents have five sensory channels.

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)
Probabilistic	Uniform	0.94	(0.94 – 0.95)	0.98	0.88	0.03
	Proportional	0.90	(0.88 – 0.91)	0.96	0.70	0.05
	Random	0.82	(0.79 – 0.84)	0.98	0.50	0.12
	Identical Random	0.92	(0.91 – 0.94)	0.99	0.79	0.05
Intelligent	Uniform	0.91	(0.89 – 0.94)	0.98	0.61	0.10
	Proportional	0.90	(0.87 – 0.93)	0.98	0.48	0.11
	Random	0.91	(0.88 – 0.93)	0.98	0.64	0.08
	Identical Random	0.90	(0.87 – 0.93)	0.98	0.45	0.11

Table 8.32: Communicative success κ — summary for agents in a *clumpy* world, with different experiences. The table shows a summary of the final range and distribution of κ across 50 repetitions of each experiment, when agents have five sensory channels.

utterance through introspective obverter, so leading to much quicker and more accurate inferences and near-optimal levels of communicative success.

8.4.3 Summary

In this section, I have investigated the effects of basing the communication experiments in a clumpy, or structured world, and compared the results to our fundamental model, in which the feature values of the objects in the world are uniformly distributed. Clearly, a structured world is much more realistic an environment for agents to inhabit than a random one, and given the completely abstract nature of the meaning creation and communication algorithms in the model, the results I obtain provide very encouraging supporting evidence that very successful communication can indeed evolve through the inference of meaning, without explicit meaning transfer, knowledge of the topic of conversation or feedback to guide the communication process. Tables 8.31 and 8.32 summarise the results for this section, and demonstrate the following findings for the experiments carried out within a clumpy world:

- under the probabilistic tree growth strategy, levels of meaning similarity σ are unchanged, but levels of communicative success κ are massively increased;
- under the intelligent tree growth strategy, levels of meaning similarity are greatly increased, leading again to yet higher levels of communicative success, even when the agents have different, random biases.
- in all cases, very high levels of communicative success are achieved as the agents exploit the structure of the information in the world.

8.5 Summary

This concludes the first detailed analysis that I have made of the introspective obverter algorithm for inferring the meaning of an utterance from its repeated use in different contexts. Making use of introspective obverter means that we can avoid the problems of explicit meaning transfer which have bedevilled other experimental simulations, and can explore agents' communicative prowess without the need for extra aid in the form of mind-reading, information about the referrent of communication, or feedback from any outside bodies about the success of the interpretative and communicative processes. Motivated by research into how children acquire the meanings of the words in their languages, I have exhaustively explored the effects of different cognitive and environmental biases on the agents' construction of conceptual structures and on their communicative success when they use the conceptual structures they have created. We have clearly seen that the conditions under which the experiments are carried out are, unsurprisingly, crucial to the results which the agents produce; the most important findings being:

- there is a strong relationship between the level of meaning similarity σ in the conceptual structures built by the agents and the communicative success κ which those agents can achieve using introspective obverter; the exact relationship varies according to the experimental conditions, but it is always a logarithmic, J-shaped curve, with the level of communicative success always higher than the level of meaning similarity;

- in a randomly-generated world, the sharing of cognitive biases between agents is very important under the probabilistic tree growth strategy, but not at all important under the intelligent tree growth strategy;

- if agents have the same experiences in the world and use the intelligent tree growth strategy, then they will build similar meaning structures and communicate more successfully, but if they build concepts probabilistically, then their experiences have no effect;

- in a structured world, both strategies have different positive effects: if the agents use probabilistic tree growth, their meaning similarity is not improved but yet they still communicate highly successfully; if they use the intelligent tree growth strategy, which takes account of the clumpiness of the world, to build their conceptual structure, then the agents will produce very high levels of meaning similarity and near-optimal levels of communicative success.

Most importantly, all these experiments show us conclusively that agents can communicate effectively with their own, individually created meanings, by inferring the meanings of words solely from their use in a variety of contexts, without the explicit transfer of meanings, without knowledge of the topic of conversation, and without feedback about the success of their learning or of the communication process in general. The inference of meaning from context is successful under many conditions, but especially so when the agents have developed their meanings intelligently, exploiting the information in the environment to produce ecologically rational conceptual representations.

CHAPTER 9

Mutual Exclusivity Revisited

> "For children to acquire vocabulary as rapidly as they do, they must be able to eliminate many potential meanings of words. One way children may do this is to assume category terms are mutually exclusive." (Markman & Wachtel, 1988, p.121)

9.1 Introduction

In chapter 8, I showed experimentally that, when using the basic introspective obverter algorithm to infer the meaning of words, there is a strong relationship between the level of co-ordination between agents' meaning structure and the level of communicative success the agents can achieve. Moreover, I found that the sharing of cognitive biases produced higher levels of meaning similarity, if the meaning creation process is driven probabilistically by those same cognitive biases; that the experiences the agents have are very important for the level of meaning similarity if the agents build meanings in an intelligent or ecologically rational way; and that this intelligent method of meaning construction is especially helpful in a clumpy world, whose structured environment it can exploit to develop near-optimal communication systems.

In chapter 3, our discussion about the possible existence of cognitive biases to explain the lexical acquisition of vocabulary items was not restricted to biases on category creation, but also included biases on interpretation. Many of these suggestions, by for instance Barrett (1986), Merriman (1986), Clark (1987) and Markman (1989), essentially boil down to the proposal that:

"children should construct mutually exclusive extensions of the terms they acquire." (Merriman & Bowman, 1989, p.1).

We investigated in particular, in chapter 3, Clark (1987)'s *Principle of Contrast*, in which a child assumes that every difference in form marks a difference in meaning, and Markman (1989)'s *Mutual Exclusivity Bias*, in which she proposes that a child assumes that the extensions of its categories are distinct sets which do not overlap; though there are differences between them in terms of both theory and explanatory emphasis, I will treat all these related proposal as different versions of an over-arching *assumption of mutual exclusivity*.

9.1.1 Mutual Exclusivity Effects

Merriman and Bowman (1989)'s analysis of the implications behind the mutual exclusivity bias, moreover, show that there are at least three different, but related, ways in which the bias could affect learning the meaning of a new word[1]:

disambiguation: if there is ambiguity in the reference of an unfamiliar word, the learner could assume it refers to the novel referent;

correction: the learner could change the extension of a familiar word in order to accommodate the introduction of the new term;

rejection: the learner could reject the new word as a synonym of an existing word;

Disambiguation of reference has been shown experimentally a number of times, particularly by Markman and Wachtel (1988), who investigated mutual exclusivity in pre-school children, and by Merriman and Bowman (1989), who compared the use of mutual exclusivity in both toddlers and pre-schoolers. Markman and Wachtel, for instance, describe their experiments in which young children were presented with random pairs of objects, one of which is familiar to them, such as a banana or a spoon, and one of which is unfamiliar, such as a lemon wedge presser or a pair of tongs. The children, on being presented with both objects, were asked by the experimenters to "show me the x", where x was a randomly chosen nonsense syllable. Markman and Wachtel found that the children are much more likely to interpret x as referring to the tongs, rather than the banana; they

[1]Merriman and Bowman also distinguish a *restriction* effect which could influence word generalisations, but it is clear that this is actually a sub-category of the *correction* effect.

hypothesise that this is because the children already understand a word which means BANANA, and they assume, under the mutual exclusivity bias, that the unfamiliar word must therefore refer to the unfamiliar object, or, as they put it: "When children hear a novel term in the presence of a familiar and unfamiliar object, children are able to use mutual exclusivity to determine the referent of the novel term." (Markman & Wachtel, 1988, p.128)

Merriman and Bowman replicated these experiments and confirmed the results obtained by Markman and Wachtel, and moreover discovered, by questioning the children about the reasons for their choices of referent, that although both groups of children appeared to use mutual exclusivity in naming new items, the older children justified their selections explicitly in terms of the mutual exclusivity principle, while the younger children did not, suggesting that children's awareness of mutual exclusivity may emerge as they develop, and may not be present from the beginning of language acquisition, as has been assumed by Clark (1987) and Markman (1989), among others.

Merriman (1986) tested the immediate correction effect by teaching children a nonsense name for a novel object. The children were then asked whether any of several other objects could be referred to with the new name, and if so, were taught a second, contrasting name for this object. Merriman found, however, that there was no apparent use of mutually exclusive extensions in this case, and no difference between children who were taught a second name and those who were not. The third potential effect suggested by Merriman, that of immediate rejection, whereby the children deny the appropriateness of a new name explicitly given to an object, or merely ignore the experimenter, has not been conclusively demonstrated, though this is not too surprising if we assume that category creation is occurring simultaneously with word learning; immediate rejection requires that the child is very confident that the categories it has created are correct and do not need to be changed.

Both immediate correction and immediate rejection, however, rely on the explicit naming of objects, and so there is no ambiguity of reference. Immediate rejection, as we have seen, also requires that concepts are stable and is at odds with immediate correction, which modifies concepts in response to conflicts of reference. The model of communication which I have been describing throughout this thesis, however, is of course based on the inference of both sense and reference through exposure in multiple contexts. Clearly, it is the disambiguation effect of mutual exclusivity which is most relevant to this model which combines concept creation and development, communication through the inference of meaning, and lexical acquisition, and therefore the disambiguation of reference will be the focus of the mutual exclusivity experiments in this chapter.

9.2 Implementing Mutual Exclusivity

In this chapter, I shall implement the mutual exclusivity bias in the model and investigate what effects its inclusion has on the development of co-ordinated meanings and successful communication. This will be done by comparing communication systems built by agents with an innate predisposition to use mutual exclusivity in disambiguating the referent of an utterance with communication systems built by agents without this predisposition; the latter experiments, of course, we have already explored in chapter 8. Two factors, in particular, are crucial in triggering the use of mutual exclusivity, and must be taken into account in developing the model; these are given below:

signal novelty: the utterance in question is novel, and unfamiliar to the learner;

disambiguation of reference through prior knowledge: the learner reduces the set of meanings under consideration by excluding all objects for which it already understands a word.

Under normal circumstances within my model, the hearer would, on hearing a word in context, build a set of all possible semantic hypotheses and use these to decipher the utterance, as I first described in section 6.5.2. The model must now be modified so that the set of semantic hypotheses is reduced by the exclusion of all referents which are already known, as described above. In addition to disambiguation of reference and the inference of the meaning of an unfamiliar word, however, Markman and Wachtel also hypothesise that mutual exclusivity can help the child to develop new meanings, when they cannot interpret an unfamiliar word, because

> "children would be left with a word for which they have not yet figured out a meaning. This should then motivate children to find a potential meaning for the novel term." (Markman & Wachtel, 1988, p.153).

The interpretation process, therefore, must be further modified to take account of Markman and Wachtel's hypothesis that interpretation failure can itself trigger the development of new conceptual structure. When the hearer encounters a new signal which it has never before encountered, therefore, its interpretation process now follows the following course:

1. it works through the objects in the context, excluding those objects for which it already knows an appropriate word. An appropriate word is defined here as a

word which the agent would use, in this context, to describe the object, and which therefore represents a meaning which would distinguish this object from all the other objects in the current context. These excluded objects, continuing the analogy with Markman and Wachtel, can be referred to as *'banana'* objects;

2. the agent is then left with a set of unfamiliar, *'tongs'* objects, and it assumes, under mutual exclusivity, that the speaker is referring to one of these objects. It therefore creates a list of semantic hypotheses based only on the *'tongs'* objects, and then interprets the word as before, choosing the meaning in which it has the highest confidence probability;

3. if no interpretation is possible, i.e. there are no appropriate meanings which distinguish any of the unfamiliar objects from all the others in the context, then the agent searches through the unfamiliar objects in turn, trying to create a new, appropriate meaning which will be appropriate to describe it in this context;

4. because this kind of meaning creation is triggered by mutual exclusivity, it proceeds by testing potential refinements on the appropriate leaf nodes of the sensory channels, until it finds a node which, once refined, will distinguish this object from all the other objects in the context, from both familiar *'banana'* objects and unfamiliar *'tongs'* objects. This method of meaning creation, although similar, is not identical to the intelligent tree growth strategy, under which sensory channels are checked to find an appropriate leaf node which could discriminate the target object from the context; in this case, the hearer does not know the target object, but has only reduced the set of possible referents to the *'tongs'* objects; it therefore checks each of these unfamiliar objects in turn until a suitable node is found;

5. the agent then creates this new meaning, and associates it with the new signal it has just encountered.

This means that there are now two potential sources of meaning creation in the model: not only failure in a discrimination game, as in all the experiments we looked at in chapter 8, but also encountering a novel signal and being unable to find an appropriate referent for it from the context. In order to explore this kind of interpretation-driven meaning creation process, it is therefore necessary to reconfigure the structure of the experiments. Instead of a two-phase process, with 1000 discrimination games during which meanings are created, followed by 5000 communication games during which meanings are not developed and the meanings of the lexical items are inferred through usage, I implement a one-phase process, with 5000 combined discrimination and communication games;

both agents' meaning creation processes continue throughout the whole experiment, but are now activated by *different* triggers, as follows:

- The speaker creates meanings as a response to failure in the discrimination game;
- The hearer creates meanings as a response to failure in the interpretation of unfamiliar words in the communication process.

This implementation of the mutual exclusivity bias differs from my earlier implementation of the principle of contrast (A. Smith 2003a). Both sets of simulations use broadly the same framework of meaning creation and communication as I describe in this thesis and the hearer's meaning creation algorithm is triggered by not being able to find any semantic hypotheses, but the main and important difference between the two experiments is that A.Smith (2003a) did not divide the objects in the context into familiar and unfamiliar sets before the hearer tried to interpret the utterance. This meant that the meaning creation process was therefore triggered only very infrequently, and so the results for the principle of contrast did not differ significantly from those without.

9.3 Mutual Exclusivity in a Random World

Without mutual exclusivity, in section 8.2, we found that there was an important difference between probabilistic and intelligent tree growth in a random world: high levels of meaning similarity under probabilistic tree growth were dependent on the sharing of cognitive biases, but this was unimportant under the intelligent strategy, where all the results were both very similar to each other, and much lower than under the probabilistic strategy. Communicative success levels were dependent on meaning similarity levels, and thus much lower under intelligent tree growth than probabilistic tree growth. The implementation of mutual exclusivity described in section 9.2 bears a number of similarities to the standard intelligent tree growth strategy, as we have seen. We might expect, therefore, that, when comparing simulations with mutual exclusivity to those without, there might be fewer differences under the intelligent tree growth strategy and greater differences under the probabilistic tree growth strategy.

Probabilistic Tree Growth

In table 9.1, we can see that the levels of meaning similarity are substantially reduced when the agents have uniform biases and the speaker uses the probabilistic tree growth

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.2)
2	0.70	(0.68 – 0.73)	0.86	0.51	0.12	0.52 **
3	0.63	(0.61 – 0.65)	0.85	0.45	0.13	0.61 **
5	0.53	(0.50 – 0.56)	0.74	0.25	0.18	0.82 **
10	0.38	(0.35 – 0.40)	0.55	0.19	0.21	0.92 **

Table 9.1: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the *probabilistic* tree growth strategy based on uniform channel biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.2, as shown in the far right-hand column.

strategy, significantly lower than the corresponding results when the experiments were divided into two distinct phases and did not assume mutual exclusivity (see table 8.2, where for instance $\sigma = 0.62$, when ten channels are available, compared to only $\sigma = 0.38$ in the current experiment). Remember that in these experiments, the speaker and hearer are now necessarily using both different triggers and different algorithms for meaning creation. This difference in trigger leads to differences in the type and amount of conceptual structure which is created by the agents.

The speaker creates a lot of conceptual structure in response to failing the discrimination games, especially early in the experiments, and this conceptual structure is dispersed throughout its sensory channels, according to its cognitive biases; the hearer, on the other hand, hears relatively few unfamiliar words, and so creates much less conceptual structure. In addition, any structure the hearer does create is more useful, because it has served to provide an appropriate meaning to the word, so there is much less redundant meaning structure.

Under the probabilistic strategy in section 8.2, the most important factor which impacted on meaning similarity was that the agents had identical biases. When mutual exclusivity is implemented, however, the hearer no longer uses its cognitive biases in meaning creation, and so there is no benefit in the agents having identical biases.

Intelligent Tree Growth

If the speaker uses the intelligent tree growth strategy, on the other hand, as in table 9.2, the levels of meaning similarity σ are now significantly *higher* than without the assumption of mutual exclusivity. Moreover, the intelligent tree growth strategy now produces

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.10)
2	0.93	(0.90 – 0.95)	1.00	0.50	0.11	0.36 **
3	0.80	(0.77 – 0.83)	1.00	0.32	0.20	0.48 **
5	0.59	(0.56 – 0.63)	0.86	0.29	0.23	0.40 **
10	0.41	(0.38 – 0.44)	0.59	0.14	0.26	0.34 **

Table 9.2: Meaning similarity σ in a random world, after agents have had 1000 different discrimination games and created individual meaning structures using the *intelligent* tree growth strategy based on uniform channel biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.10, as shown in the far right-hand column.

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)
Probabilistic	Uniform	0.53	(0.50 – 0.56)	0.74	0.25	0.18
	Proportional	0.41	(0.37 – 0.45)	0.72	0.15	0.36
	Random	0.41	(0.37 – 0.44)	0.67	0.00	0.34
	Identical Random	0.48	(0.45 – 0.51)	0.78	0.28	0.22
Intelligent	Uniform	0.59	(0.56 – 0.63)	0.86	0.29	0.23
	Proportional	0.53	(0.49 – 0.58)	0.88	0.20	0.30
	Random	0.50	(0.45 – 0.55)	0.90	0.15	0.36
	Identical Random	0.55	(0.50 – 0.60)	0.90	0.18	0.31

Table 9.3: Meaning similarity σ — summary for agents in a random world, with different experiences. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of σ across 50 repetitions of each experiment, when agents have five sensory channels.

higher levels of meaning similarity than the probabilistic tree growth strategy, in a reversal of the results in chapter 8. The exclusion of familiar objects from the meaning inference process, which of course helps the disambiguation process considerably, focuses the hearer's meaning creation, and this may account for the rise in meaning similarity, although it is quite probable that the extended timescale in these experiments, where meaning creation can potentially take place at any point, also has an effect.

Although tables 9.1 and 9.2 refer only to the default scenario where the agents have uniform biases, if we compare the summary table 9.3 to table 8.14, we find that the effects of both the cognitive biases and the tree growth strategies have been largely neutralised by the implementation of mutual exclusivity. Specifically, we can see that:

- meaning similarity rates have decreased considerably when agents have identical biases and use the probabilistic tree growth strategy; when they have different random biases, there is no change;
- meaning similarity rates have increased under the intelligent tree growth strategy, so that these are now higher than the probabilistic results;
- cognitive biases now have very little effect on the level of meaning similarity, although uniform biases always produce higher results than random biases, because there is necessarily less variation in the configuration of the agents' biases between different runs of the same experiment.

Communicative Success

Given the strong relationship between meaning similarity and communicative success which has been a feature of this model, we might expect that we would see a drop in communicative success under the probabilistic tree growth strategy, and an increase in communicative success under the intelligent tree growth strategy. Although this is indeed what we see in table 9.4, where communicative success has fallen significantly under the probabilistic tree growth strategy, there is actually no change under the intelligent tree growth strategy, except when only two channels are available. Moreover, if we compare all the results in table 9.6, we find that there are no major differences in the level of communicative success at all; all the various permutations of tree growth strategy and cognitive bias produce average levels of communicative success in the same range $(0.67 \leq \kappa \leq 0.76)$.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.3)
2	0.88	(0.86 – 0.90)	0.96	0.64	0.08	0.68 **
3	0.80	(0.77 – 0.82)	0.94	0.49	0.11	0.84 **
5	0.70	(0.67 – 0.72)	0.88	0.46	0.12	0.94 **
10	0.56	(0.54 – 0.58)	0.75	0.39	0.14	0.95 **

Table 9.4: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the *probabilistic* tree growth strategy based on uniform biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.3, as shown in the far right-hand column.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.11)
2	0.91	(0.89 – 0.93)	0.99	0.57	0.08	0.35 **
3	0.82	(0.80 – 0.85)	0.98	0.54	0.14	0.21
5	0.73	(0.70 – 0.76)	0.96	0.42	0.15	0.24
10	0.58	(0.55 – 0.61)	0.78	0.34	0.17	0.14

Table 9.5: Communicative success κ in a random world, after 5000 communicative episodes following meaning creation using the *intelligent* tree growth strategy based on uniform biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in table 8.11, as shown in the far right-hand column.

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)
Probabilistic	Uniform	0.70	(0.67 – 0.72)	0.88	0.46	0.12
	Proportional	0.72	(0.68 – 0.75)	0.92	0.49	0.16
	Random	0.69	(0.66 – 0.72)	0.93	0.33	0.17
	Identical Random	0.76	(0.73 – 0.78)	0.91	0.49	0.12
Intelligent	Uniform	0.73	(0.70 – 0.76)	0.96	0.42	0.15
	Proportional	0.74	(0.71 – 0.77)	0.93	0.48	0.15
	Random	0.67	(0.64 – 0.70)	0.89	0.41	0.18
	Identical Random	0.69	(0.66 – 0.73)	0.95	0.42	0.17

Table 9.6: Communicative success κ — summary for agents in a random world, with different experiences. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of κ across 50 repetitions of each experiment, when agents have five sensory channels.

Full details of all these experiments and others not explicitly reported can be found in appendix D, but it is clear that our expectations over the relative changes of the two different tree growth strategies have been borne out: meaning similarity is much lower under the probabilistic strategy, and slightly higher under the intelligent strategy.

In a random world, moreover, the assumption of mutual exclusivity in the disambiguation process of interpretation, and the triggering of meaning creation when this fails, effectively irons out many of the distinctions which we found in chapter 8, with regard to the levels of communicative success which the agents achieve; this is no longer affected very much either by the agents' cognitive biases nor the speaker's tree growth strategy.

9.4 Mutual Exclusivity in a Clumpy World

In a clumpy world without mutual exclusivity and interpretation-driven meaning creation, which I described in section 8.4, we found that very high, near-optimal average levels of communicative success were achieved, as the agents exploited the structure of the information in their environment, despite relatively low levels of meaning similarity, particularly under the probabilistic tree growth strategy. Similar levels of communicative success were achieved under all cognitive biases, though levels of meaning similarity were much higher under the intelligent strategy than the probabilistic strategy (see table 8.31), as this strategy is much more able to take account of the structure of the world.

Probabilistic Tree Growth

When we introduce mutual exclusivity into a similar set of simulations, we find similar results to those we saw in section 9.3, but the effect on meaning similarity is even more pronounced; in table 9.7, we can see that average levels of meaning similarity $\bar{\sigma}$ are extremely low indeed under the probabilistic tree growth strategy, significantly lower than without the assumption of mutual exclusivity. This decrease is particularly noticeable as the number of channels available increases, but in all cases, the value of the KS statistics ($KS = 1$) implies that the two sets of distributions could not be more completely different from each other. Again, this happens because the agents are using completely different strategies for meaning creation, which have more pronounced effects in a clumpy world. The speaker's strategy, namely meaning creation in response to discrimination failure, is essentially unaffected by the structure of the world, but is controlled by its cognitive biases; the hearer's, on the other hand, is driven by trying to interpret meaning from context, and is very much affected by the structure of the world, which will lead it

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.23)	KS(9.1)
2	0.50	(0.47 – 0.52)	0.66	0.26	0.18	1.0 **	0.74 **
3	0.43	(0.41 – 0.46)	0.62	0.25	0.21	1.0 **	0.73 **
5	0.35	(0.33 – 0.37)	0.55	0.24	0.17	1.0 **	0.82 **
10	0.21	(0.20 – 0.22)	0.31	0.14	0.17	1.0 **	0.85 **

Table 9.7: Meaning similarity σ in a *clumpy* world, after agents have had 1000 different discrimination games and created individual meaning structures using the *probabilistic* tree growth strategy based on uniform channel biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity.* The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.23 and 9.1, as shown in the far right-hand columns.

Channels	Mean $\bar{\sigma}$	CI	Max(σ)	Min(σ)	CoV(σ)	KS(8.27)	KS(9.2)
2	0.95	(0.91 – 0.99)	1.00	0.42	0.14	0.28 *	0.48 **
3	0.97	(0.95 – 0.99)	1.00	0.58	0.08	0.38 **	0.75 **
5	0.92	(0.88 – 0.95)	1.00	0.60	0.12	0.32 **	0.80 **
10	0.85	(0.82 – 0.89)	1.00	0.63	0.13	0.32 **	0.98 **

Table 9.8: Meaning similarity σ in a *clumpy* world, after agents have had 1000 different discrimination games and created individual meaning structures using the *intelligent* tree growth strategy based on uniform channel biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity.* The table shows a summary of the final range and distribution of σ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.27 and 9.2, as shown in the far right-hand columns.

to build structure on those channels on which objects can be distinguished, and not on those where there are clumps of objects as described in section 8.4.

Interestingly, however, when we look at the summary table 9.9, we can see that average levels of meaning similarity with proportional biases produce higher levels, although still relatively low in comparison with other experiments. Proportional biases, of course, are designed to reflect the structure of a clumpy world to a larger extent than random or uniform biases, and higher levels of meaning similarity result from both agents exploiting the structure of the world in different ways.

Intelligent Tree Growth

Under the intelligent tree growth strategy, on the other hand, meaning similarity levels are always subject to very large increases, with significantly higher levels of $\bar{\sigma}$ in comparison

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)
Probabilistic	Uniform	0.35	(0.33 – 0.37)	0.55	0.24	0.17
	Proportional	0.52	(0.49 – 0.55)	0.78	0.23	0.22
	Random	0.34	(0.31 – 0.38)	0.74	0.05	0.35
	Identical Random	0.37	(0.34 – 0.40)	0.59	0.14	0.31
Intelligent	Uniform	0.92	(0.88 – 0.95)	1.00	0.60	0.12
	Proportional	0.88	(0.84 – 0.91)	1.00	0.49	0.16
	Random	0.91	(0.87 – 0.94)	1.00	0.55	0.13
	Identical Random	0.92	(0.89 – 0.95)	1.00	0.60	0.11

Table 9.9: Meaning similarity σ — summary for agents in a *clumpy* world, with different experiences. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of σ across 50 repetitions of each experiment, when agents have five sensory channels.

with experiments without mutual exclusivity. Moreover, these experiments also produce significantly higher levels of $\bar{\sigma}$ than the experiments in a random world which we looked at in section 9.3.

Although the meaning creation algorithms for each role are still triggered differently, the combination of a clumpy, structured world and intelligent tree growth in the speaker still produces high levels of meaning similarity. Both agents are now exploiting the structure of the world in deciding where to create conceptual structure, even though discrimination is the motive for the speaker and interpretation the motive for the hearer. Of course, the structure of the world is the same for both agents, therefore they will both tend to make the same kind of distinctions in the world.

As long as the assumption of mutual exclusivity is included, then, consistently high levels of meaning similarity are found when combining intelligent tree growth with all the different cognitive bias allocations, which themselves now appear to play hardly any role in determining the relative conceptual structures built by the agents. Overall, table 9.9 shows that there is now a marked difference between the two tree growth strategies in terms of meaning similarity, and we can see that:

- in a clumpy world, high levels of meaning similarity can only arise if the speaker uses an intelligent tree growth strategy, and complete synchronisation of conceptual structure is very much achievable;

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.25)	KS(9.4)
2	0.83	(0.80 – 0.85)	0.94	0.49	0.11	0.94 **	0.32 **
3	0.83	(0.81 – 0.85)	0.95	0.60	0.08	0.88 **	0.22
5	0.81	(0.79 – 0.83)	0.93	0.57	0.09	0.90 **	0.62 **
10	0.80	(0.78 – 0.82)	0.92	0.58	0.09	0.80 **	0.87 **

Table 9.10: Communicative success κ in a *clumpy* world, after 5000 communicative episodes following meaning creation using the *probabilistic* tree growth strategy based on uniform biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.25 and 9.4, as shown in the far right-hand columns.

Channels	Mean $\bar{\kappa}$	CI	Max(κ)	Min(κ)	CoV(κ)	KS(8.29)	KS(9.5)
2	0.94	(0.92 – 0.96)	0.99	0.59	0.08	0.26	0.34 **
3	0.93	(0.91 – 0.95)	0.99	0.68	0.07	0.36 **	0.56 **
5	0.90	(0.88 – 0.92)	0.99	0.67	0.08	0.36 **	0.72 **
10	0.86	(0.84 – 0.88)	0.97	0.57	0.10	0.24	0.94 **

Table 9.11: Communicative success κ in a *clumpy* world, after 5000 communicative episodes following meaning creation using the *intelligent* tree growth strategy based on uniform biases. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity*. The table shows a summary of the final range and distribution of κ across 50 repetitions of the experiment. The distributions were compared statistically with those shown in tables 8.29 and 9.5, as shown in the far right-hand columns.

- the particular cognitive bias mechanisms have very little effect on meaning similarity.

Communicative Success

In table 9.10 we can see that despite the very low levels of meaning similarity, yet again the agents communicate surprisingly well under the probabilistic strategy. When the speaker uses the intelligent strategy, shown in table 9.11, we can again see very high levels of meaning similarity, some of which are significantly different to those obtained without mutual exclusivity, others of which are the same. Again, the intelligent strategy produces slightly higher levels of communicative success, as we have seen throughout the experiments, but the differences between the two strategies shown in table 9.12 are very small indeed compared to the differences in meaning similarity seen in table 9.9.

Tree Growth Strategy	Biases	Mean		Range		
		$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)
Probabilistic	Uniform	0.81	(0.79 – 0.83)	0.93	0.57	0.09
	Proportional	0.82	(0.80 – 0.83)	0.92	0.62	0.08
	Random	0.78	(0.75 – 0.80)	0.91	0.57	0.11
	Identical Random	0.82	(0.79 – 0.84)	0.96	0.60	0.09
Intelligent	Uniform	0.90	(0.88 – 0.92)	0.99	0.67	0.08
	Proportional	0.87	(0.85 – 0.89)	0.97	0.65	0.09
	Random	0.89	(0.87 – 0.91)	0.98	0.71	0.08
	Identical Random	0.90	(0.88 – 0.92)	0.98	0.69	0.07

Table 9.12: Communicative success κ — summary for agents in a *clumpy* world, with different experiences. The hearer's meaning creation is driven by *interpretation failure and the assumption of mutual exclusivity.* The table shows a summary of the final range and distribution of κ across 50 repetitions of each experiment, when agents have five sensory channels.

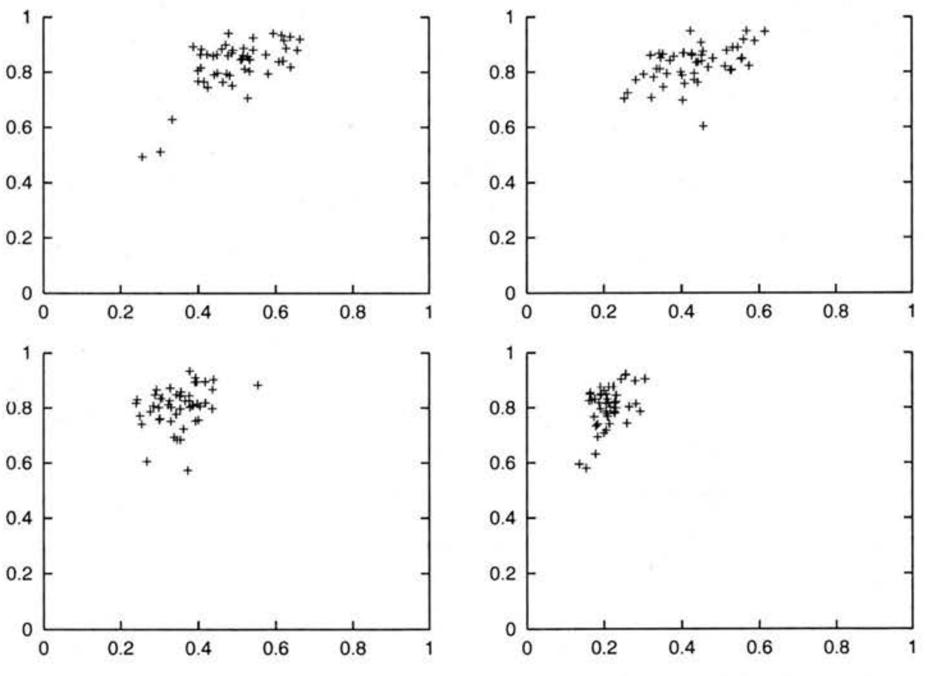


Figure 9.1: Meaning similarity σ (x-axis) against communicative success κ (y-axis) in a *clumpy* world, after 5000 combined discrimination and communicative episodes; the speaker's meaning creation is driven by discrimination game failure, using the probabilistic tree growth strategy based on *uniform* channel biases, the hearer's by *communication failure and the assumption of mutual exclusivity.* The simulation is repeated 50 times, with each run represented by a separate cross on the graph. Sub-figures show varying numbers of channels available to the agents: 2 (upper left), 3 (upper right), 5 (lower left) and 10 (lower right).

The relationship between meaning similarity and communicative success which has been so consistent throughout the experiments in chapter 8 appears now to be remarkably different, with very high levels of κ no matter what the level of meaning similarity; figure 9.1, which shows uniform biases and the probabilistic tree growth strategy, is a good example of this, with the crosses very bunched up towards the top left-hand corner of the figures rather than clustering along the $x = y$ diagonal, particularly when the agents have many channels available.

It is clear that, although the level of meaning similarity in these experiments is very dependent on the particular tree growth strategy chosen by the speaker, if we concentrate on the communicative aspects of the model, the assumption of mutual exclusivity alone allows agents to communicate very successfully in a clumpy world, without any need to have very much conceptual structure in common at all. As we have seen throughout this chapter, the agents create meaning structure in response to different pressures, which leads to predictable differences in meaning similarity depending on the environment and on the speaker's tree growth strategy. In all cases, however, the hearer is driven to create meanings only when they will enable the disambiguation of an unfamiliar word. The meanings the hearer creates are characterised by two features:

- they are useful, in that they can be used to discriminate at least some objects
- they are relatively general, because there is relatively little tree growth on the hearer's sensory channels.

We already know from the Gricean nature of the agents (see section 7.5) that the agents are likely to use meanings which are as general as possible to describe situations, and we also know that hearers using mutual exclusivity to trigger concept growth do not, in general, build massive conceptual structures. Even though the two agents have different concepts, then, the pressures on communication will lead them to use the meanings which the other is most likely to have, thus leading to communicative success which far outstrip meaning similarity on which the basic obverter model is based, as we can see in figure 9.1. Mutual exclusivity, therefore, promotes only relevant conceptual growth in the hearer, and ensures that, even if divergent semantic structures are built, the meanings which agents have in common are those which are used in communication.

9.5 Summary

In this chapter, I have developed the model further to include an assumption of mutual exclusivity into the hearer's interpretation process, both in terms of excluding familiar objects as possible referents of the unfamiliar utterance, and also of triggering meaning creation when no possible meanings are found to map to the unfamiliar utterance. This introduction of mutual exclusivity was motivated by many accounts of children's word learning, particularly those proposed by Barrett (1986), Merriman (1986), Clark (1987) and Markman (1989), but, importantly, nothing else in the model has changed, so the fundamental assumptions of avoiding explicit meaning transfer, mind-reading and feedback are still intact.

The introduction of the mutual exclusivity assumption to the hearer's interactions with the world has had some very interesting effects, the most important of which are:

- the different triggers and mechanisms of meaning creation used by the agents often result in very low levels of meaning similarity, particularly in a clumpy world when the speaker uses the probabilistic tree growth strategy, which does not exploit the information in the world;
- despite these low levels of meaning similarity σ, the agents can communicate very successfully, with levels of κ at around 70% in a randomly-generated world, and 90% in a structured, clumpy world. Levels of communicative success, indeed, are determined only by the environment and the speaker's tree growth strategy, not by its cognitive biases;

In summary, the assumption of mutual exclusivity is a very powerful addition to the model, which means that agents no longer have to have synchronised meaning structures in order to communicate successfully. Without innate or explicitly transferable meanings, without being able to read the minds of their interlocutors, without receiving any feedback about the communication process, the agents still build a successful communication system. The introduction of a mutual exclusivity assumption into the hearer's interpretation process leads to the development of fewer, but more relevant meanings in the hearer's conceptual structure, and therefore to relatively high levels of communicative success despite conceptual divergence.

CHAPTER 10

Conclusions

> "Communication does not begin when someone makes a sign, but when someone interprets another's behaviour as a sign." (Burling, 2000, p.30)

In this thesis, my principal aim has been to explore the construction of communication systems, in populations of simulated agents, based on the inference of meaning from context. This focus on the inference of meaning is based on the recognition that communication should be primarily viewed from the hearer's point of view, as Burling (2000) points out above. Communication can only occur when a hearer is available to receive a signal and to interpret its meaning, rather than when a signaller produces a signal, even if the signal is produced with the intention of conveying a meaning. Not only can communication systems not start without an interpreter, but their complex development is constrained by the speed of the development of the hearer's interpretative capabilities. Increases in signalling power are of no use whatsoever if the signals cannot be interpreted, but on the other hand increases in interpretative power *can* be useful, because, in an inferential model, the speaker's internal meaning representation and the hearer's internal meaning representation do not need to match, as we saw in section 6.3.2.

The burgeoning field of language evolution contains many recent contributions in the forms of computational simulations, which provide an ideal environment for the rigorous testing of the complex and dynamic systems which are at the heart of human language, yet few experimenters pay any attention to the models of meaning representation and meaning creation which their simulations assume. In contrast to this general trend, my work in this thesis has been underpinned by the following assumptions:

- agents do not have innate meanings, but instead construct their own semantic representation of their environment;
- the meanings of signals are not explicitly transferred during the communication process, but are instead inferred by agents from the contexts in which the signals are spoken;
- agents cannot read the minds of their interlocutors, nor do they have access to any of their interlocutors' internal mental processes;
- agents are not provided with feedback about the results of their communication, and are therefore not guided towards a communicative goal.

In the remainder of this chapter, I will outline the main findings of the work described in the thesis, before moving on to discuss future promising directions for research using inferential models of language and communication.

10.1 Summary

10.1.1 Empirical Meaning Creation

I showed in chapter 4 that although many computational models of the evolution of language have an explicitly named 'semantics' or 'meaning space', these often had very little to do with even a vaguely realistic semantic model. The majority of the models I investigated, indeed, contained categories and semantic predicates which are innate and pre-specified by the experimenters themselves, and which, more damningly, do not refer to anything in a simulated external world. Perhaps just as surprisingly, few internal sense relationships between the meanings were found either; instead the concepts were regularly atomic and isolate, bearing no resemblance to each other, and simply appearing and disappearing at the whim of the experimenter, rather than under the control of the agents. We saw that the main reason for the inclusion of 'semantics' within these models, in fact, was simply to act as a template for the agents, so that they are able to generalise across signal-meaning pairs and thereby appear to develop a more powerful 'syntax' which parallels the pre-specified 'semantics'.

In order to avoid such pitfalls, and to develop a communicative model based truly on the empirical creation and interpretation of meanings, I looked first at the nature of meaning itself, in particular and at categorisation as the most basic form of meaning, by considering, in chapter 2, various different models of categorisation which agents can use to independently build their own individual conceptual structures.

In chapter 5, I described a model of empirical meaning creation based on the *discrimination game* introduced by Steels (1996b), in which an agent must find a category which describes one particular object, and distinguishes it thus from another set of objects. Following on from this, I simulated a simple model Steelsian world containing a number of objects, each of which can be described in terms of the values of their observable features. Simulated agents interact with the objects in the world using *sensory channels*; they have the same number of sensory channels as the objects have features, and there is a one-to-one mapping between them. Sensory channels are sensitive to the objects' feature values; specifically, they can detect whether a particular feature value falls between two bounds on a sensory channel. The process of meaning creation itself takes place through *refinement*, or the splitting of a channel's sensitivity range into two discrete segments of equal size. This results in the formation of two new categories, each of which is sensitive to half the original range. Each category is itself a candidate for further refinement, so producing, over time, a hierarchical, dendritic structure, with the nodes on the tree representing categories, or *meanings* (Steels, 1999). This conceptual structure represents sense relationships through its dendritic structure, and reference relationships through its empirical grounding in the agents' external environment.

Adaptation of an agent's conceptual structure, and therefore the creation of meaning, is triggered by failure in a discrimination game. Each agent has a *tree growth strategy* for choosing a channel for refinement, which is based on its cognitive biases and/or the details of the particular discrimination game which failed, as I described in chapter 8. This flexibility in the meaning creation process allows different agents, however, to create very different conceptual structures, each of which will nevertheless be able to distinguish objects in the world; in order to compare these, I designed two similarity measures τ and σ, which allow quantifiable comparisons of agents' conceptual structures to be made.

10.1.2 Signal Redundancy and Inferential Communication

In chapter 6, we saw that any idealisation of communication which reduces semantic representations to a simple template against which a coding scheme can be constructed, and which must therefore assume the explicit transfer of meanings in conjunction with signals, leads damagingly to the design problem which I have called the *signal redundancy paradox*, which is repeated in summary form below:

- if meanings are transferable, then signals are redundant;
- but if signals are removed, then to what extent does the model represent communication?

For communication to occur, the agents must be able to decipher the utterances, and learn to associate particular meanings with particular signals, despite not being provided with the meaning which the speaker intended to convey. I assume that the most general source for the inference of meaning is the environment in which the agent is placed, and this in turn suggests that at least some of the meanings which agents talk about are likely to be used to refer to objects and events which actually happen in the environment. The existence of an external world from which meaning can be inferred is crucial to a realistic model of meaning, for without it, any 'meanings' are necessarily abstract and pre-defined, and realistic communication cannot emerge. If the meanings do not have reference, they can only be 'communicated' through explicit transfer, which of course entails the signal redundancy paradox. In order to avoid this, therefore, as I described in section 6.2, there must be at least three levels of representation in the model:

1. an external environment, which is public and accessible to all, which provides the motivation and source for meaning creation;
2. a private, agent-specific internal representation of meaning, which is not perceptible to others;
3. a set of signals, which can be transmitted between agents and is in principle public.

Thinking of the communicative function of language as a simple coding system between signals and meanings, however, is problematic not just in terms of the communication model itself, but also in terms of the evolution of such a system. It is important to remember, therefore, that language is necessarily both reciprocal and cultural. There is no communicative advantage in a *single* mutant obtaining a language acquisition device if other individuals do not have one, as the communication process is by definition interactive, and must contain at least two individual agents. Neither, however, is there any advantage in *many* mutants having a language acquisition device, while there is no language existing in the community for them to acquire. Explanations of the emergence of an LAD, therefore, must also explain the emergence of the first linguistic communication, the Bickertonian "magic moment" (Bickerton, 1990); how did the hearer of the first signal know that the signal was meaningful and was conveying a meaning? Origgi and

Sperber (2000), however, point out that mutations which allows individuals to *infer* the meanings of signals can not only provide an explanation for how language got started, but can also provide a plausible account of the progressive complication of language over evolutionary time.

At the beginning of this chapter, I discussed Burling (2000)'s insight into the instantiation of communication, which could only begin when a signal was interpreted as conveying a meaning, notwithstanding the fact that such a meaning might never have been intended by the signaller. In terms of linguistic evolutionary development, when the system is already functioning to some extent, Origgi and Sperber (2000) discuss a possible genetic mutation which allows the construction of a more complex semantic representation; if communication is not based on inference, the explicit transfer of more complex, incompatible semantic structure from one agent to another will cause confusion and communicative breakdown, and, being communicatively harmful, such a mutation is unlikely to maintain itself in the population. In an inferential model, on the other hand, the mismatch between the speaker's meaning and the hearer's meaning does not have catastrophic effects on communication, because individuals can have very different internal representations of meanings, and yet can still communicate successfully. Those without the new mutation, for instance, who therefore still possess only the more basic semantic representation, could still communicate with other individuals in blissful ignorance of a more complex semantic structure, while the mutants with the enhanced semantic representation might receive an additional advantage in terms of more accurate or detailed inference of the meaning. Because of their advantage, they might, in time, develop new ways of representing the patterns they accidentally find in this structure. This insight is enshrined in my model through the use of reference identity to evaluate communication.

In chapter 6, I explored Hurford (1989)'s findings that *lexical bidirectionality* is very important in the evolution of optimal communication systems. One such communicative algorithm, which explicitly encodes lexical bidirectionality is the obverter algorithm described by Oliphant and Batali (1997). Unfortunately, however, this algorithm required the agents to have access to each other's internal mental representations, thus violating one of the design goals of this work. Mindful of this, and of the primacy of interpretation over production, I modified the obverter procedure so that the speaker's production behaviour was itself based on interpretation. In my model, therefore, the speaker chooses a signal by first putting itself in the hearer's shoes, and choosing a signal which it would understand, if it heard the signal in this same situation, and had to infer its meaning from context. This modified, *introspective obverter* methodology allows agents to communicate without explicit meaning transfer, so avoiding the signal redundancy paradox, without knowledge of the topic of conversation, and without feedback about the communicative process itself.

10.1.3 Meaning Similarity and Communicative Success

In chapter 7, I demonstrated through a detailed series of experiments that there is a strong correlation between the level of meaning similarity σ and communicative success κ in this model. If meanings are allocated randomly, communicative success κ is regularly higher than meaning similarity σ, because the hierarchical meaning creation process exerts pressure in favour of balanced tree structures, and the meanings which are most likely to be shared by agents are also those for which the agents prefer to use in communication, in accordance with Gricean conversational maxims.

Learning the meanings of words, of course, is utterly unremarkable to children, who effortlessly overcome Quine (1960)'s problem of indeterminacy. In view of this, in chapter 3, I examined proposals concerning the existence of constraints within the learners themselves which predispose them to disregard some of the theoretically possible meanings of a signal, thus reducing the size of the set of semantic hypotheses, and making Quine's problem soluble. In chapter 8, I explored the effects of different cognitive and environmental biases on the agents' construction of conceptual structures and on their communicative success thereafter, and found that the relationship between meaning similarity σ and communicative success κ remains strong. In a randomly-generated world, the agents cannot improve on creating meanings based on their cognitive biases, using a *probabilistic* tree growth strategy; high levels of conceptual similarity will always arise if the agents share similar values of these biases. In a structured, or clumpy world, on the other hand, then it is much better for the agents to use a more *intelligent*, ecologically rational (Gigerenzer & Todd, 1999) tree growth strategy, which can exploit the information in the environmental structure to a much greater degree.

Motivated furthermore by psychologists' suggestions of interpretation biases which help children learn vocabulary, I then implemented Markman (1989)'s *mutual exclusivity assumption*, by which an individual uses prior knowledge to help disambiguate the reference of novel signals, a process which affects both the interpretation of utterances and the building of new conceptual structure. In chapter 9, I show experimentally that agents no longer need to have synchronised meaning structures in order to communicate successfully. The introduction of a mutual exclusivity assumption into the hearer's interpretation process leads to the development of fewer, but more relevant meanings in the hearer's conceptual structure, and therefore to relatively high levels of communicative success despite conceptual divergence.

10.2 Future Developments

The model of meaning creation and communication described throughout this thesis has allowed me to present a number of important results, the most important of which is that successful communication systems can be constructed by independent, unguided agents, through the inference of meaning. It is tempting to conclude from this that, because communication can arise without explicit meaning transfer, we can go back to building simplified models which incorporate innate and transferable meanings. I believe however, that such a conclusion would result in missing an important opportunity for further explanatory research, which is motivated by the realisation that the characteristics of communication systems built on inferred meanings could *in themselves* explain universals of human language, both in terms of structural properties and dynamic processes, which cannot easily be explained by appealing to an Chomskyan innate language acquisition device alone.

In this respect, it is important to note that, although communication through the inference of meaning is very successful, especially in a structured world when ecologically rational strategies are followed, it is rarely if ever perfect. We can easily see that these imperfections in the communication process, which are not imposed by adding noise to the model, but which arise simply through the inferential mechanisms through which communication systems are developed, lead inexorably to variation in the agents' languages and their conceptual structures. Human languages are by nature dynamic, and just as inferential communication leads to synchronic variation, the very same synchronic variation drives all historical language change (Trask, 1996). One profitable avenue of research in this framework will certainly be the detailed investigation of different kinds of structural linguistic change, including such processes as *grammaticalisation* (Hopper & Traugott, 1993), in which more complex grammatical markers such as case markers and complementisers are created from less complex lexical items over generations of inference, which occur directly as a result of the dynamic and imperfect nature of the communication process. Indeed, the process of grammaticalisation itself has been explicitly described by leading researchers as "context-induced reinterpretation" (Heine & Kuteva, 2002, p.3), and it is clear that an inferential model of language such as that described here directly parallels this view of grammatical change, providing an ideal framework for the exploration of grammaticalisation processes in particular, and language change in general.

Models such as the *iterated learning model* (ILM) described by Hurford (2002), in which language agents are situated in populations, and their knowledge is transmitted culturally from adults to children, have already shown that social and population pressures can lead in themselves to the emergence of linguistic structure (Kirby, 2001; Brighton, 2002; Brighton et al., 2003). Vogt and Coumans (2003), indeed, have already shown, in a replication of some of the basic communicative simulations in this thesis, that the dynamics of the ILM in themselves provide a boost to the time taken for communication success to occur, and it seems reasonable to assume that such benefits might also be maintained in larger populations, although this remains an open question for the time being.

I have shown in this thesis, therefore, that successful communication can emerge and evolve through the repeated inference of meaning from context. In the longer term, I anticipate that the overarching aim of research in this framework will be to explore whether both the structural properties of language which have evolved over generations of use, and the processes of language change itself, can themselves be explained as emergent properties of the repeated cycle of signal production and the inference of meaning. I give the final word on the potential value of this explanatory paradigm to the respected linguist Ray Jackendoff:

> "If some aspects of linguistic behaviour can be predicted from more general considerations of the dynamics of communication in a community, rather than from the linguistic capabilities of individual speakers, then they should be." (Jackendoff, 2002, p.101)

APPENDIX A

Model Outline

The model of meaning creation and communication which is used in this thesis can appear relatively complex, containing many different stages and procedures. This appendix contains a brief description of the important parts of the model for reference.

1. Initialisation.

   The world is initialised, and the following items are created according to the parameters of the experiment:

   - Agents are provided with
     - sensory channels with empty discrimination trees on which they will build conceptual structure, one channel for each feature with which the objects are described (see below);
     - sensory channel biases for each of their sensory channels (see section 8.2.1);
     - possibly innate meanings (see section 7.2);
     - empty lexicons of signal-meaning pairs (see section 6.5).
   - Objects are defined in terms of feature values:
     - these can be *randomly* distributed through the feature value space
     - or *clumped* together in groups (see section 8.4).

2. Interactive Episodes.

   Each episode consists of an obligatory discrimination episode, possibly followed by a communicative episode.

(a) Discrimination (see chapter 5):

- Two agents are selected from the population and assigned to the roles of *speaker* and *hearer*.
- A set of objects is constructed, and is called the *context*.
- One of these objects is chosen to be the *target object*.
- Both context and target are provided to the speaker, which searches through its conceptual structure to try to find a meaning which describes the target and does not describe the other objects in the context.
- If the speaker finds a suitable meaning, called the *speaker's meaning*, then the game succeeds, and the episode moves on to *communication*.
- If the speaker fails to find any suitable meaning, the game fails:
  - the speaker chooses a sensory channel according to its tree growth strategy (see section 8.2.2), and refines the node on this channel corresponding to the target object;
  - the episode now ends, without the communicative episode taking place.

(b) Communication (see chapter 6):

- The speaker chooses an utterance to express the speaker's meaning it found in the discrimination game, using the *introspective obverter* algorithm (see section 6.5.1);
- The utterance and the context (but not the details of which object is the target object) are provided to the hearer;
- The hearer compiles a list of *possible referents*:
  - in the standard model, this list is simply all the objects in the context;
  - if the agent is being guided by the *mutual exclusivity assumption* (see chapter 9), however, and the utterance is novel, then the hearer excludes all objects in the context for which it already has an appropriate word, and is then left with a set of unfamiliar objects as *possible referents* (see section 9.2).
- The hearer goes through the list of possible referents, and plays a separate discrimination game for each of them, with the possible referent as the target object.
- Each of these games produces zero or more meanings which could be used to distinguish the temporary target object from the other objects in the context;

- Collating these meanings results in a list of possible meanings or *semantic hypotheses* which the hearer considers;
- If there is at least one semantic hypothesis:
  - The hearer associates, in its lexicon, all the semantic hypotheses with the utterance;
  - From the semantic hypotheses, it chooses the *hearer's meaning*, which is defined as the meaning in which it has the most confidence (see section 6.5.2);
- If the list of possible meanings is empty, then a new meaning is created;
  - If the agent is being guided by the *mutual exclusivity assumption* (see chapter 9), then the meaning is created according to the procedure in section 9.2;
  - Otherwise, the agent chooses a sensory channel according to its tree growth strategy (see section 8.2.2), and refines the node on this channel corresponding to an object chosen at random from the context;
- The hearer associates this new meaning with the utterance, in its lexicon, and the new meaning becomes the *hearer's meaning*;
- The object in the context which is referred to by the hearer's meaning, the *hearer's referent*, is compared to the target object (see section 6.3.2):
  - if they match, the communicative episode succeeds;
  - if they do not match, the communicative episode fails.

APPENDIX B

Experimental Measures

Measure	Symbol	Description
Coefficient of Variation	CoV(x)	The standard deviation of x expressed as a percentage of the mean $\bar{x}$ (see section 7.3).
Cognitive Bias	b_{a_c}	The cognitive bias on agent a's sensory channel c (see section 8.2).
Communicative Success Rate	κ	The percentage of successful communication episodes (see section 6.3).
Confidence Probability	$p(s, m)$	The conditional probability that, given a particular signal s, the meaning m can be expected (see equation 6.1).
Depth	$d(\lambda)$	The depth in the discrimination tree of node λ (see chapter 7.2).
Discriminative Success Rate	δ	The percentage of successful discrimination games (see section 5.2.1).
Discriminative Success Probability	$P_{\Delta(c)}$	The probability of discrimination game Δ succeeding because a distinctive category is found on sensory channel c (see equation 5.12).

Measure	Symbol	Description
Groups	$g(c)$	In the definition of a clumpy world, the number of groups on a sensory channel c (see section 8.4).
Kolmogorov-Smirnov Statistic	KS	A comparative measure of whether two sample distributions are drawn from the same population distribution function (see section 8.1.1).
Meaning Similarity	$\sigma(a_1, a_2)$	The meaning similarity between two agents a_1 and a_2 (see equation 5.15).
Number of Different Possible Trees	$\phi(n)$	The number of different discrimination trees which can be created with n refinements (see equation 7.2).
Proportional Bias Constant	p	The constant on which the allocation of proportional biases is based (see equation 8.1).
Refined Leaf Node	$\lambda(c)$	The leaf node on sensory channel c which is refined in the meaning creation process (see 8.2.2).
Tree Similarity	$\tau(t_1, t_2)$	The similarity between two trees t_1 and t_2 (see equation 5.13).
Unique Discriminability	ψ	The percentage of objects in the model which can be distinguished from *all* other objects in the world (see section 5.2.3).
Usage	$u(s, m)$	The number of communicative episodes in which signal s has been associated with meaning m (see section 6.5).

Table B.1: Reference table of experimental measures

APPENDIX C

Experimental Results

This appendix contains comprehensive results for the experiments reported in chapter 8, where levels of meaning similarity and communicative success are calculated under the following different parameters:

Tree Growth Strategy: the strategy used by the agents in creating meanings following discrimination failure;

Cognitive Biases: the method of bias allocation for the agents' biases;

Structure of the World: the method of constructing the agents' environment;

Agents' Experiences: whether the agents have the same or different interactions with their environment;

Hearer's Meaning Creation: in appendix C, the hearer's meaning creation is driven by discrimination failure in all experiments.

For each experiment, the same summary figures and tables are shown:

- at the top right, a scatter plot shows the relationship between meaning similarity σ and communicative success κ at the end of each simulation run;
- down the left, line plots show the progression of meaning similarity σ (upper), and of communicative success κ (lower) over time; each simulation run is shown with a separate line;
- to the right of these plots, the tables summarise the final results obtained, showing the average, range and variation in the values of σ and κ respectively.

Each figure itself consists of four sub-figures, which show results for experiments conducted with different numbers of sensory channels/features, as follows:

upper left objects are defined by *two* feature values, and agents have two corresponding sensory channels;

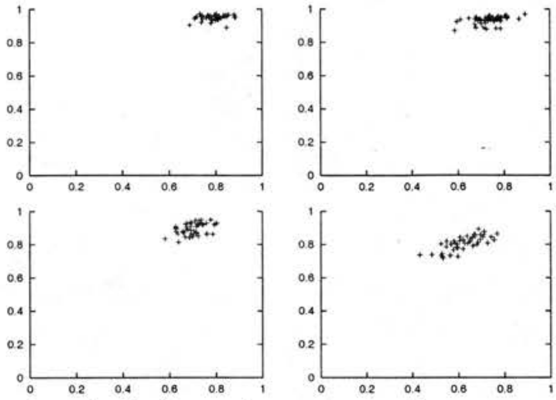
upper right objects are defined by *three* feature values, and agents have three corresponding sensory channels;

lower left objects are defined by *five* feature values, and agents have five corresponding sensory channels;

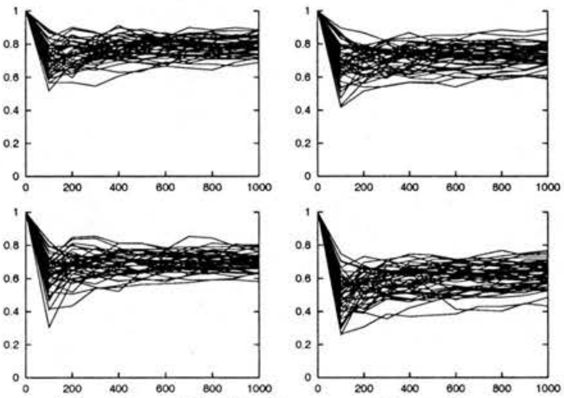
lower right objects are defined by *ten* feature values, and agents have ten corresponding sensory channels.

C.1 Random World

Tree Growth Strategy:	**probabilistic**
Biases:	**uniform**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



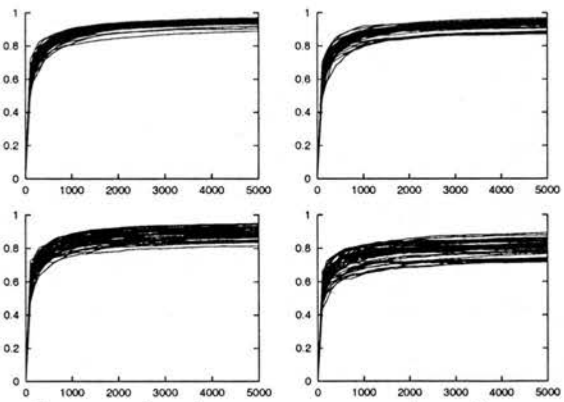
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)
2	0.80	(0.78 – 0.81)	0.89	0.69	0.06
3	0.74	(0.72 – 0.76)	0.89	0.59	0.08
5	0.70	(0.69 – 0.71)	0.80	0.58	0.07
10	0.62	(0.61 – 0.64)	0.77	0.43	0.11

Summary of the final values of σ.



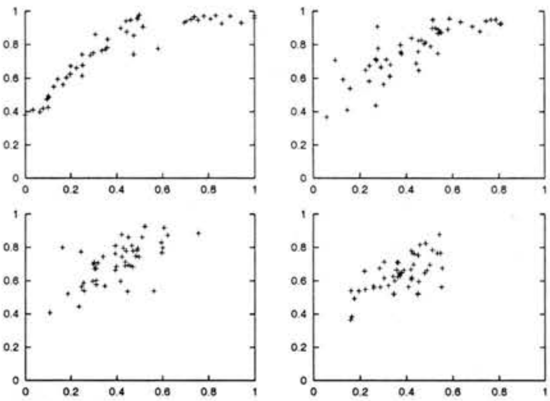
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)
2	0.95	(0.94 – 0.95)	0.97	0.89	0.02
3	0.93	(0.93 – 0.94)	0.97	0.87	0.02
5	0.90	(0.89 – 0.91)	0.95	0.82	0.04
10	0.81	(0.80 – 0.82)	0.89	0.72	0.05

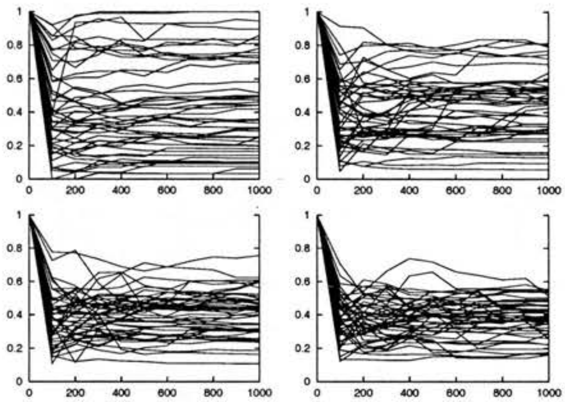
Summary of the final values of κ.

Figure C.1: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



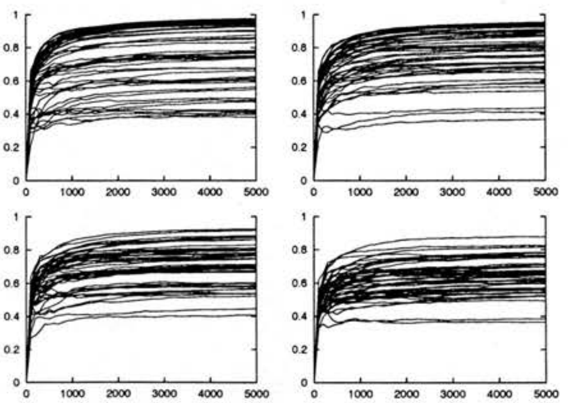
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.1)
2	0.42	(0.34 – 0.50)	1.00	0.00	0.66	0.76 **
3	0.44	(0.38 – 0.49)	0.81	0.06	0.44	0.82 **
5	0.41	(0.37 – 0.44)	0.76	0.11	0.32	0.96 **
10	0.38	(0.35 – 0.41)	0.55	0.16	0.29	0.86 **

Summary of the final values of σ.



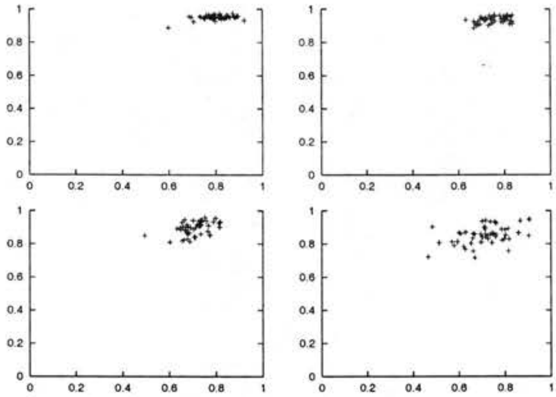
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.1)
2	0.76	(0.71 – 0.82)	0.98	0.38	0.25	0.62 **
3	0.77	(0.73 – 0.81)	0.95	0.37	0.19	0.70 **
5	0.71	(0.68 – 0.74)	0.93	0.41	0.17	0.84 **
10	0.64	(0.61 – 0.67)	0.88	0.37	0.16	0.80 **

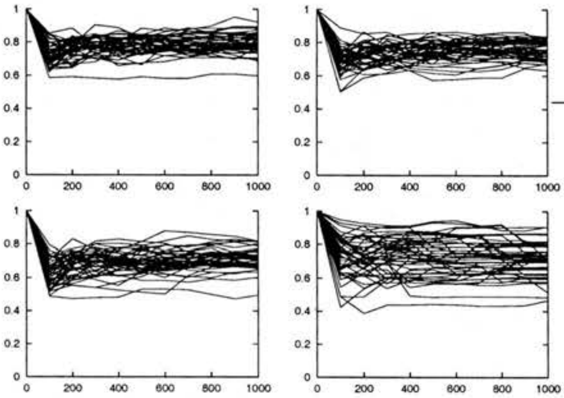
Summary of the final values of κ.

Figure C.2: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy: **probabilistic**
Biases: **proportional** ($p = 0.5$)
World: **random**
Experiences: **different**
Hearer's Concept Creation driven by: **discrimination**



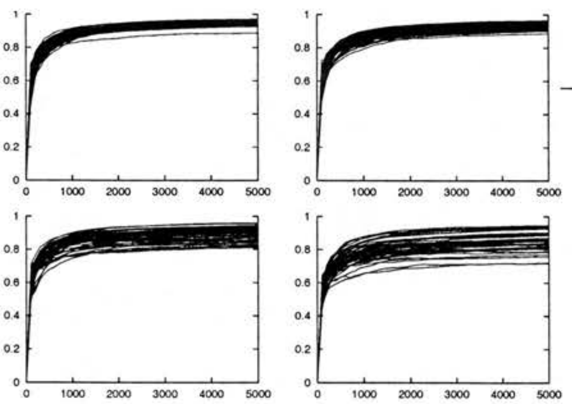
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.1)	KS (C.2)
2	0.80	(0.78 – 0.82)	0.92	0.60	0.07	0.12	0.74 **
3	0.76	(0.74 – 0.77)	0.84	0.63	0.07	0.24	0.84 **
5	0.71	(0.69 – 0.73)	0.82	0.50	0.08	0.20	0.94 **
10	0.71	(0.68 – 0.74)	0.91	0.47	0.14	0.50 **	0.92 **

Summary of the final values of σ.



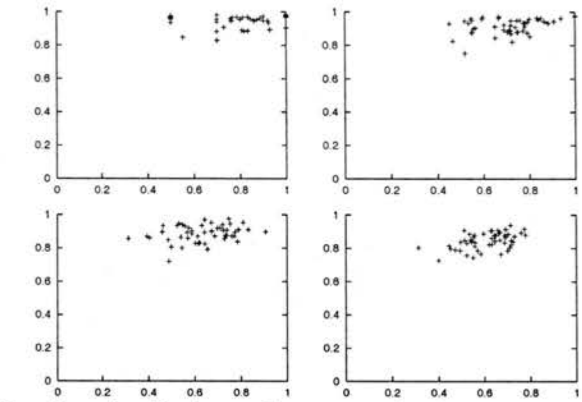
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.1)	KS (C.2)
2	0.95	(0.95 – 0.96)	0.97	0.89	0.01	0.12	0.64 **
3	0.93	(0.93 – 0.94)	0.97	0.89	0.02	0.14	0.78 **
5	0.90	(0.89 – 0.91)	0.96	0.81	0.04	0.16	0.82 **
10	0.85	(0.83 – 0.87)	0.95	0.72	0.06	0.42 **	0.82 **

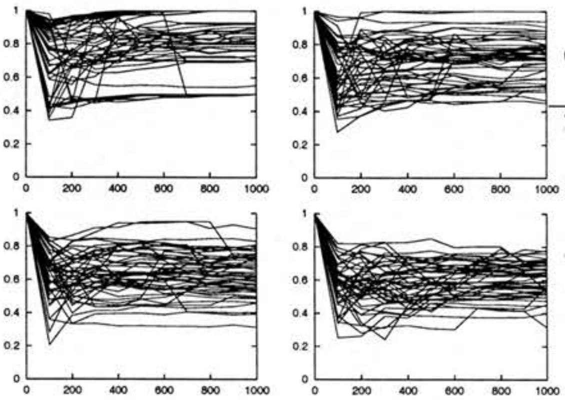
Summary of the final values of κ.

Figure C.3: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**identical random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



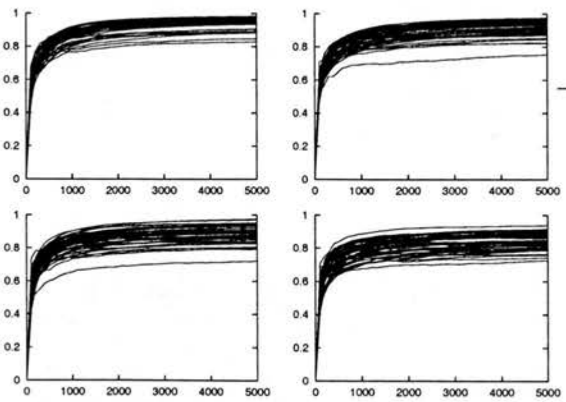
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.1)	KS (C.2)	KS (C.3)
2	0.81	(0.76 – 0.85)	1.00	0.50	0.21	0.36 **	0.68 **	0.34 **
3	0.71	(0.68 – 0.75)	1.00	0.45	0.18	0.26	0.66 **	0.32 **
5	0.63	(0.60 – 0.67)	0.91	0.31	0.19	0.44 **	0.70 **	0.48 **
10	0.61	(0.58 – 0.64)	0.78	0.31	0.17	0.20	0.76 **	0.44 **

Summary of the final values of σ.



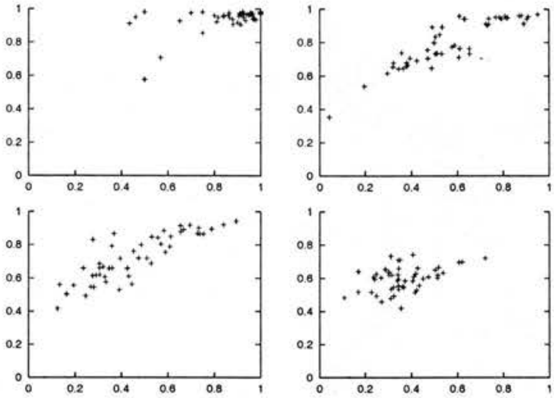
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.1)	KS (C.2)	KS (C.3)
2	0.95	(0.94 – 0.96)	0.98	0.83	0.04	0.48 **	0.56 **	0.44 **
3	0.92	(0.90 – 0.93)	0.98	0.75	0.05	0.30 *	0.58 **	0.34 **
5	0.89	(0.87 – 0.90)	0.97	0.72	0.06	0.14	0.76 **	0.14
10	0.85	(0.83 – 0.86)	0.94	0.73	0.06	0.38 **	0.84 **	0.12

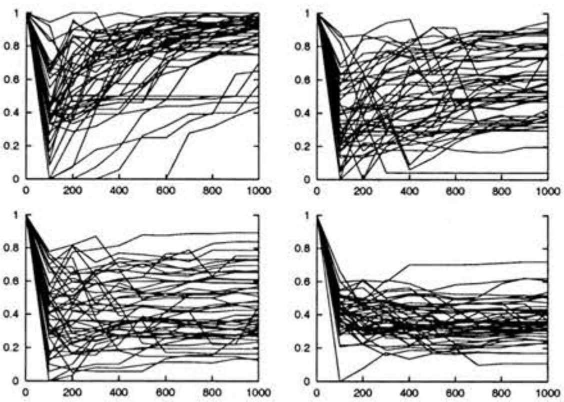
Summary of the final values of κ.

Figure C.4: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**uniform**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



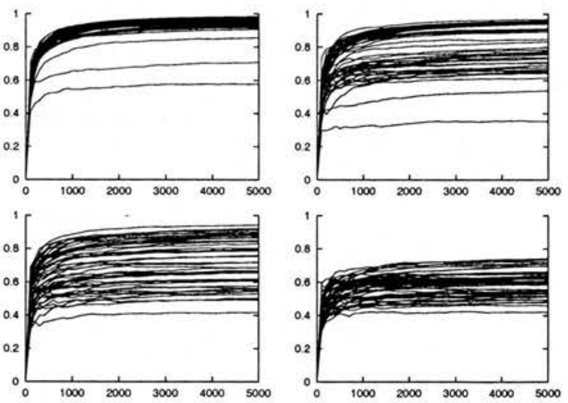
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.1)
2	0.86	(0.82 – 0.91)	1.00	0.43	0.17	0.64 **
3	0.58	(0.53 – 0.64)	0.95	0.04	0.35	0.58 **
5	0.46	(0.41 – 0.51)	0.89	0.12	0.42	0.76 **
10	0.37	(0.33 – 0.40)	0.72	0.11	0.32	0.88 **

Summary of the final values of σ.



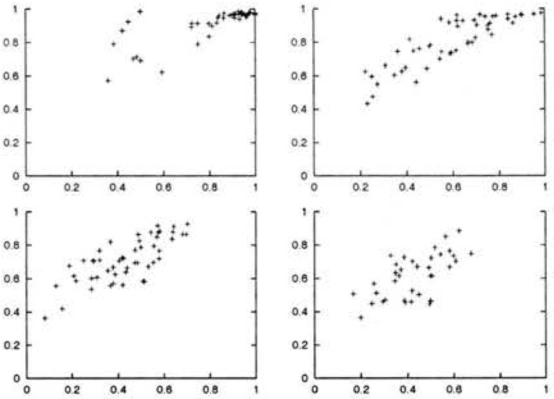
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.1)
2	0.94	(0.92 – 0.96)	0.98	0.58	0.07	0.44 **
3	0.80	(0.76 – 0.84)	0.97	0.35	0.17	0.56 **
5	0.72	(0.68 – 0.76)	0.94	0.42	0.19	0.66 **
10	0.60	(0.57 – 0.62)	0.74	0.42	0.12	0.94 **

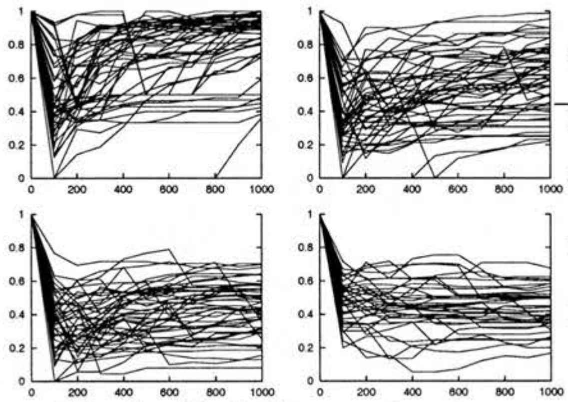
Summary of the final values of κ.

Figure C.5: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



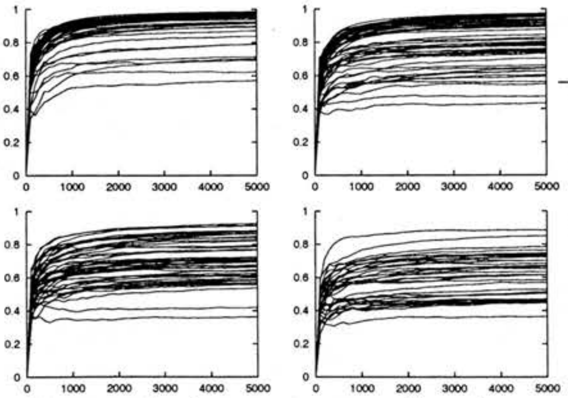
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.2)	KS (C.5)
2	0.82	(0.76 – 0.87)	1.00	0.36	0.23	0.58 **	0.14
3	0.60	(0.55 – 0.66)	0.99	0.22	0.34	0.42 **	0.14
5	0.43	(0.39 – 0.48)	0.70	0.08	0.35	0.22	0.18
10	0.43	(0.39 – 0.47)	0.68	0.17	0.29	0.29 *	0.32 *

Summary of the final values of σ.



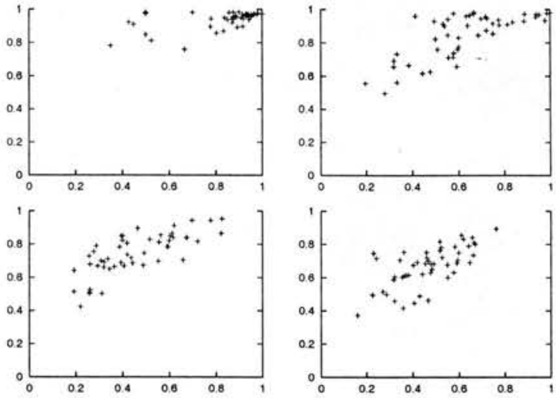
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.2)	KS (C.5)
2	0.92	(0.89 – 0.94)	0.99	0.57	0.11	0.44 **	0.18
3	0.81	(0.77 – 0.85)	0.97	0.43	0.18	0.22	0.14
5	0.71	(0.68 – 0.75)	0.93	0.36	0.18	0.12	0.12
10	0.61	(0.57 – 0.65)	0.89	0.36	0.21	0.29 *	0.29 *

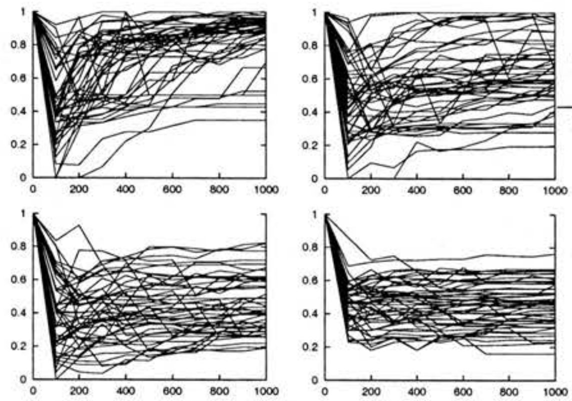
Summary of the final values of κ.

Figure C.6: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**proportional** ($p = 0.5$)
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



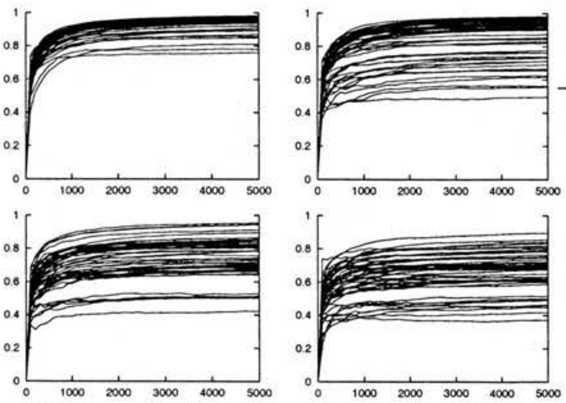
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.3)	KS (C.5)	KS (C.6)
2	0.84	(0.79 – 0.88)	1.00	0.35	0.20	0.56 **	0.18	0.12
3	0.63	(0.57 – 0.68)	1.00	0.20	0.32	0.60 **	0.18	0.080
5	0.46	(0.41 – 0.51)	0.83	0.19	0.37	0.80 **	0.080	0.18
10	0.47	(0.43 – 0.50)	0.76	0.16	0.29	0.70 **	0.40 **	0.16

Summary of the final values of σ.



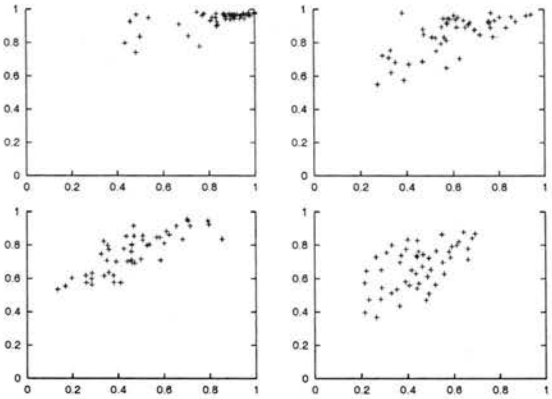
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.3)	KS (C.5)	KS (C.6)
2	0.94	(0.92 – 0.95)	0.98	0.76	0.05	0.34 **	0.14	0.10
3	0.84	(0.80 – 0.88)	0.98	0.50	0.16	0.46 **	0.18	0.16
5	0.74	(0.71 – 0.78)	0.95	0.42	0.17	0.68 **	0.26	0.18
10	0.67	(0.63 – 0.70)	0.89	0.37	0.18	0.74 **	0.46 **	0.25

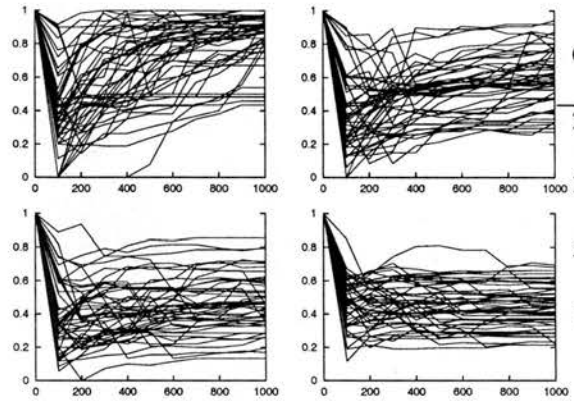
Summary of the final values of κ.

Figure C.7: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**identical random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



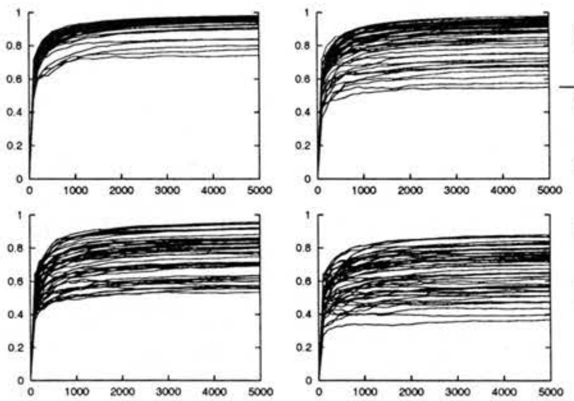
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.5)	KS (C.6)	KS (C.7)
2	0.84	(0.80 – 0.88)	1.00	0.43	0.18	0.20	0.12	0.14
3	0.60	(0.56 – 0.65)	0.94	0.27	0.28	0.18	0.10	0.12
5	0.47	(0.42 – 0.51)	0.85	0.13	0.35	0.16	0.12	0.14
10	0.44	(0.41 – 0.48)	0.69	0.21	0.29	0.38 **	0.11	0.14

Summary of the final values of σ.



Communicative success κ against time.

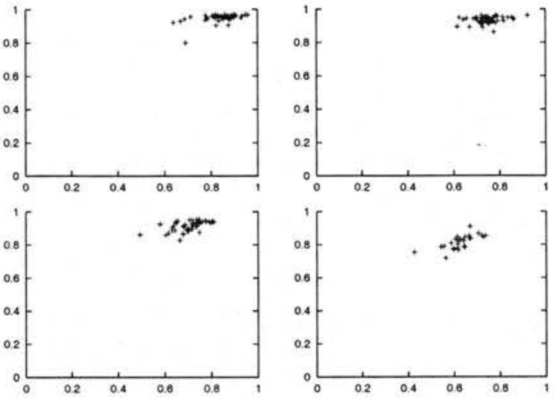
Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.5)	KS (C.6)	KS (C.7)
2	0.94	(0.93 – 0.96)	0.98	0.74	0.06	0.14	0.16	0.10
3	0.85	(0.82 – 0.88)	0.98	0.55	0.13	0.26	0.24	0.14
5	0.75	(0.72 – 0.79)	0.96	0.53	0.16	0.22	0.20	0.14
10	0.67	(0.63 – 0.70)	0.88	0.37	0.20	0.42 **	0.22	0.14

Summary of the final values of κ.

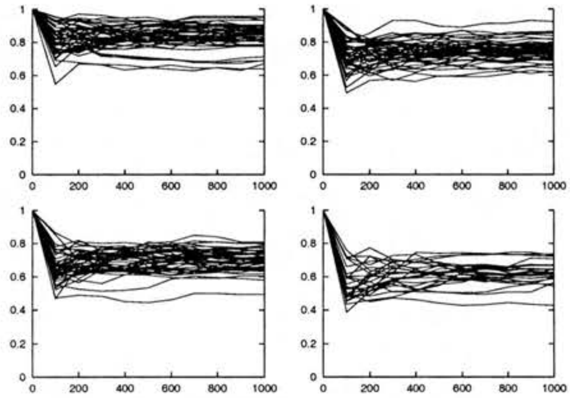
Figure C.8: Meaning similarity σ, communicative success κ (see box for parameters).

C.2 Identical Experiences

Tree Growth Strategy:	**probabilistic**
Biases:	**uniform**
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



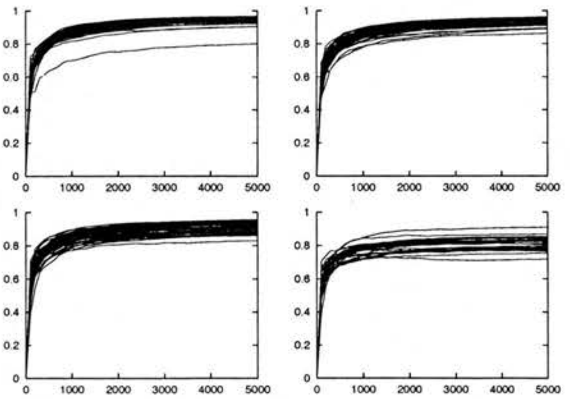
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.1)
2	0.84	(0.82 – 0.86)	0.96	0.64	0.08	0.50 **
3	0.75	(0.73 – 0.77)	0.92	0.62	0.08	0.14
5	0.70	(0.69 – 0.72)	0.81	0.49	0.09	0.18
10	0.63	(0.60 – 0.65)	0.74	0.43	0.10	0.17

Summary of the final values of σ.



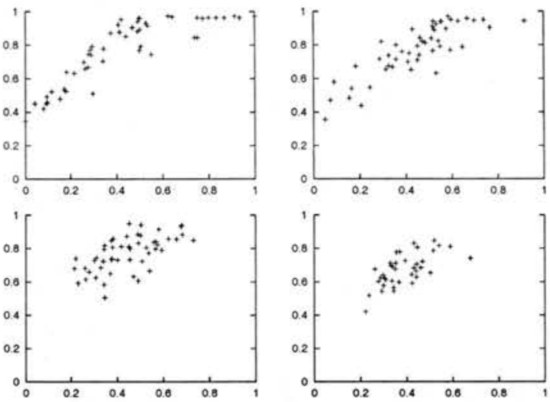
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.1)
2	0.95	(0.94 – 0.96)	0.97	0.80	0.03	0.14
3	0.93	(0.93 – 0.94)	0.96	0.86	0.02	0.10
5	0.91	(0.91 – 0.92)	0.96	0.83	0.03	0.30 *
10	0.81	(0.80 – 0.83)	0.91	0.72	0.05	0.14

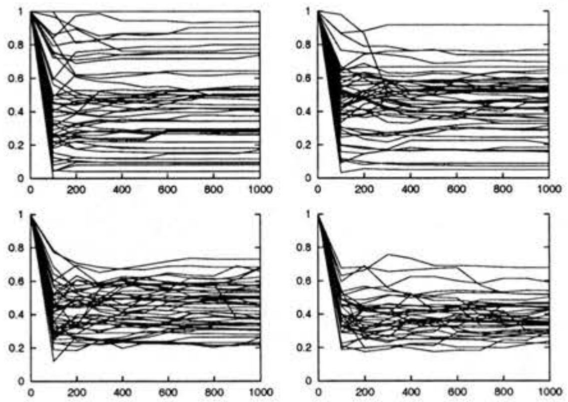
Summary of the final values of κ.

Figure C.9: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**random**
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



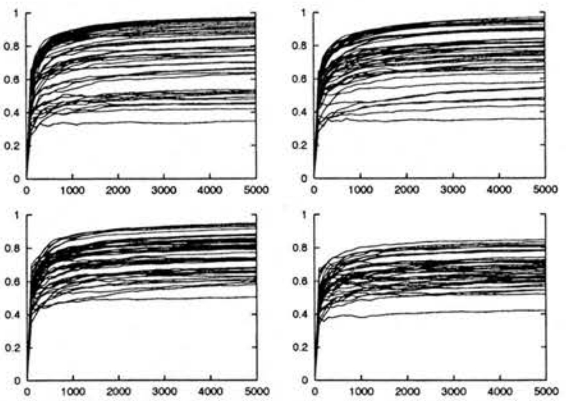
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.2)
2	0.43	(0.36 – 0.50)	1.00	0.00	0.59	0.14
3	0.45	(0.40 – 0.50)	0.92	0.05	0.41	0.12
5	0.45	(0.41 – 0.48)	0.73	0.22	0.29	0.18
10	0.39	(0.36 – 0.42)	0.68	0.22	0.25	0.14

Summary of the final values of σ.



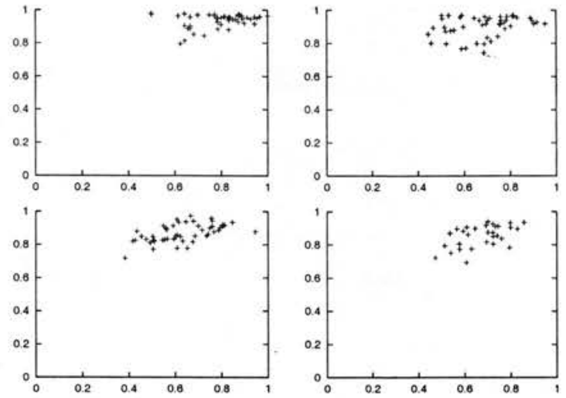
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.2)
2	0.77	(0.72 – 0.82)	0.97	0.35	0.24	0.10
3	0.77	(0.73 – 0.81)	0.97	0.36	0.20	0.100
5	0.77	(0.74 – 0.80)	0.95	0.50	0.14	0.28 *
10	0.67	(0.64 – 0.70)	0.85	0.42	0.14	0.20

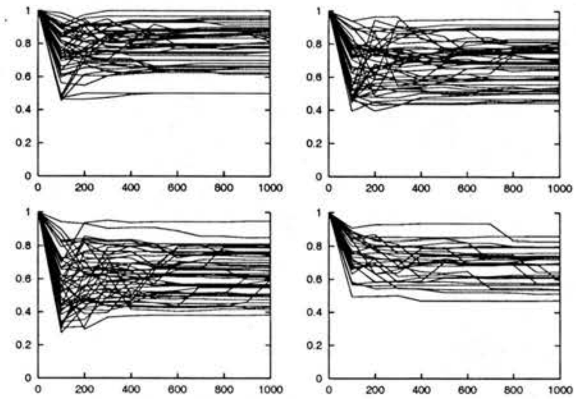
Summary of the final values of κ.

Figure C.10: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**proportional** ($p = 0.5$)
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



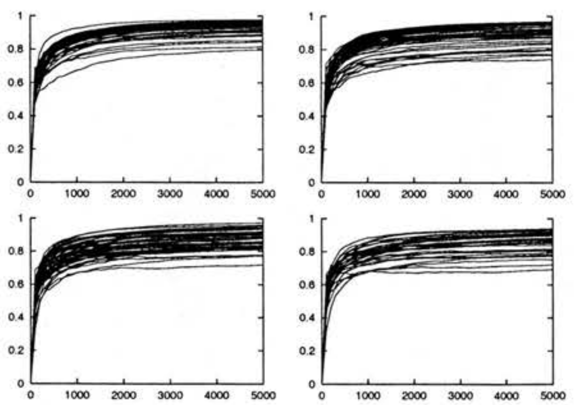
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.3)
2	0.79	(0.76 – 0.82)	1.00	0.50	0.15	0.22
3	0.69	(0.66 – 0.73)	0.95	0.44	0.19	0.36 **
5	0.64	(0.60 – 0.67)	0.95	0.38	0.20	0.50 **
10	0.67	(0.64 – 0.71)	0.86	0.47	0.15	0.21

Summary of the final values of σ.



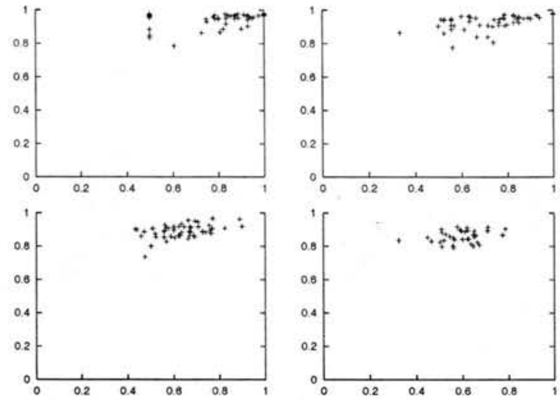
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.3)
2	0.94	(0.92 – 0.95)	0.98	0.80	0.04	0.24
3	0.90	(0.89 – 0.92)	0.97	0.74	0.07	0.34 **
5	0.87	(0.85 – 0.88)	0.97	0.72	0.06	0.30 *
10	0.85	(0.83 – 0.88)	0.94	0.69	0.08	0.23

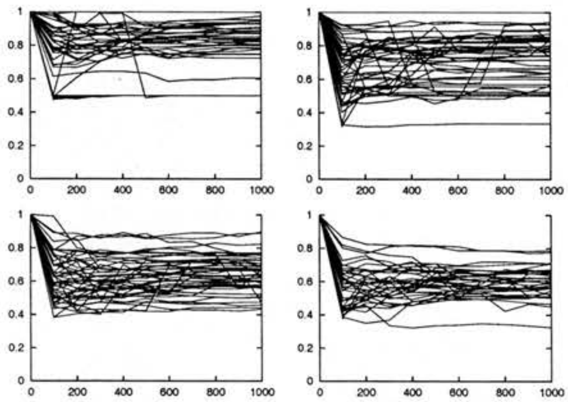
Summary of the final values of κ.

Figure C.11: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**identical random**
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



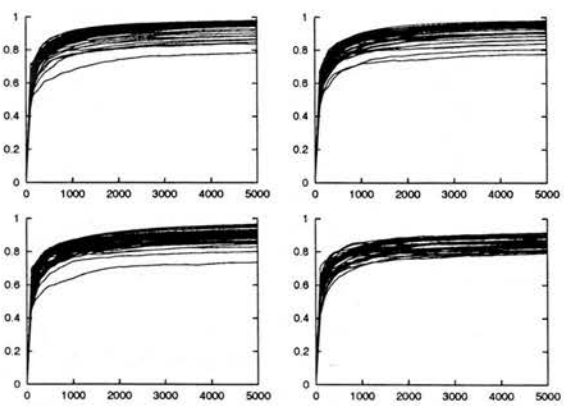
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.4)
2	0.79	(0.74 – 0.84)	1.00	0.50	0.22	0.14
3	0.72	(0.68 – 0.77)	1.00	0.33	0.20	0.14
5	0.64	(0.61 – 0.67)	0.90	0.43	0.17	0.12
10	0.60	(0.57 – 0.63)	0.79	0.32	0.15	0.20

Summary of the final values of σ.



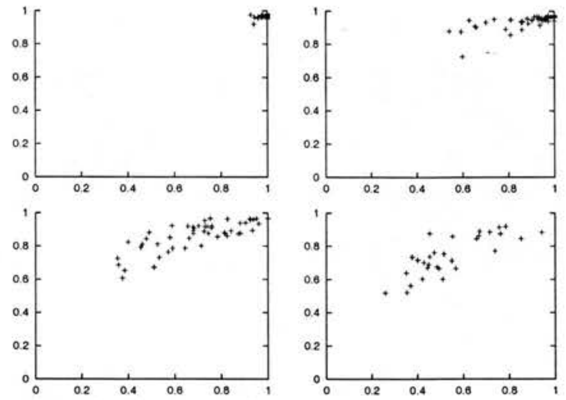
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.4)
2	0.94	(0.93 – 0.95)	0.98	0.79	0.04	0.22
3	0.93	(0.91 – 0.94)	0.98	0.78	0.05	0.20
5	0.89	(0.88 – 0.90)	0.97	0.74	0.05	0.12
10	0.86	(0.85 – 0.87)	0.92	0.79	0.04	0.18

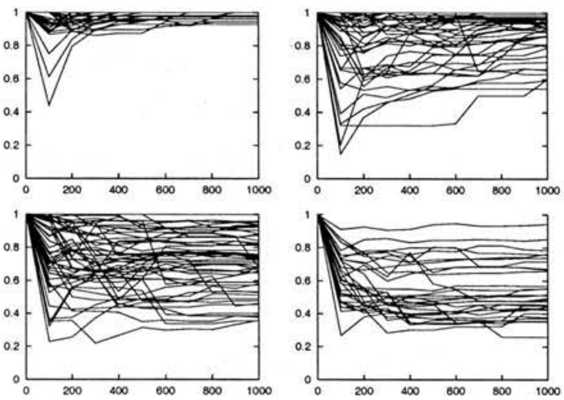
Summary of the final values of κ.

Figure C.12: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**uniform**
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



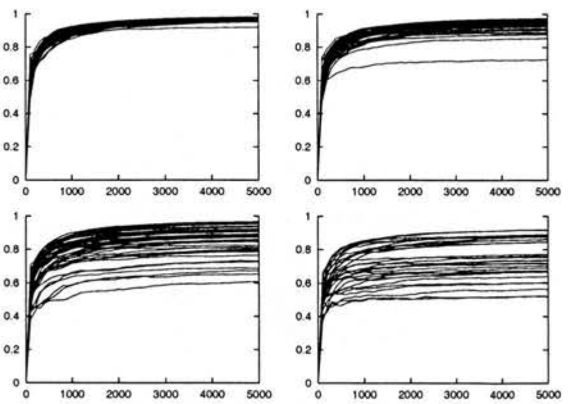
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.5)
2	0.99	(0.99 – 1.00)	1.00	0.93	0.02	0.74 **
3	0.89	(0.86 – 0.93)	1.00	0.54	0.14	0.66 **
5	0.70	(0.65 – 0.75)	1.00	0.35	0.25	0.48 **
10	0.55	(0.49 – 0.61)	0.94	0.26	0.30	0.55 **

Summary of the final values of σ.



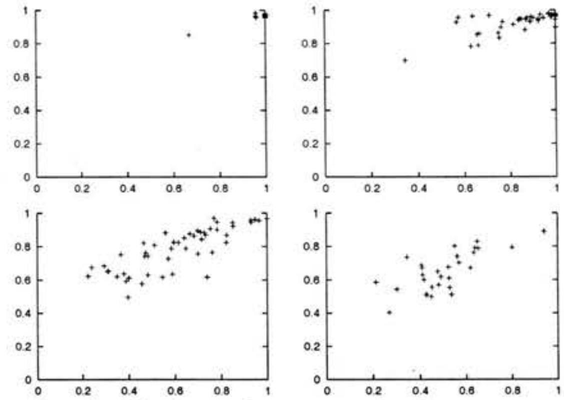
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.5)
2	0.97	(0.97 – 0.97)	0.98	0.92	0.01	0.36 **
3	0.94	(0.93 – 0.95)	0.98	0.73	0.04	0.54 **
5	0.86	(0.84 – 0.89)	0.97	0.61	0.10	0.46 **
10	0.75	(0.70 – 0.79)	0.92	0.52	0.16	0.66 **

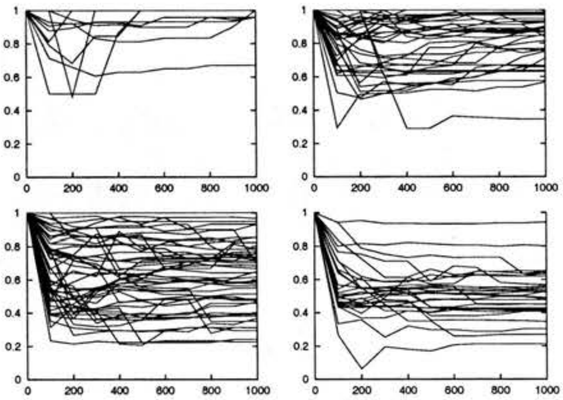
Summary of the final values of κ.

Figure C.13: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**random**
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



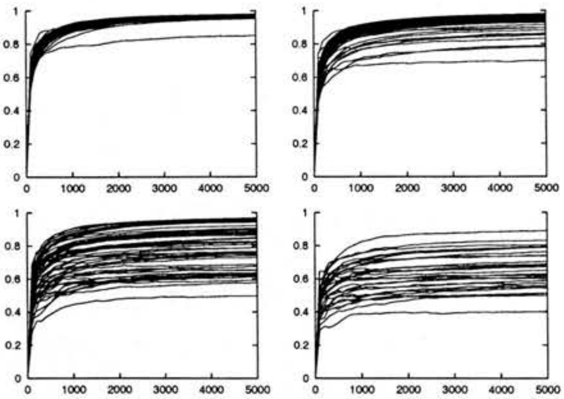
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.6)
2	0.99	(0.98 – 1.00)	1.00	0.67	0.05	0.82 **
3	0.87	(0.83 – 0.91)	1.00	0.35	0.17	0.58 **
5	0.61	(0.56 – 0.67)	1.00	0.23	0.33	0.46 **
10	0.51	(0.45 – 0.56)	0.94	0.21	0.29	0.29

Summary of the final values of σ.



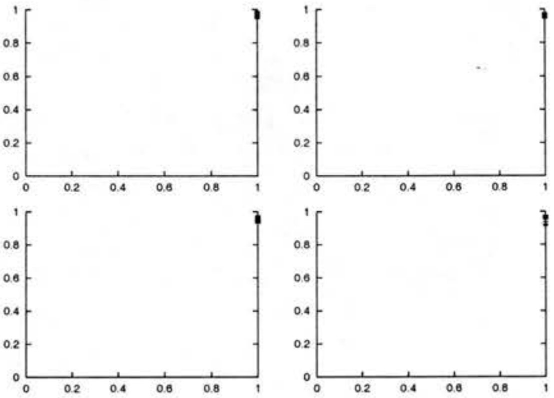
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.6)
2	0.97	(0.96 – 0.97)	0.98	0.85	0.02	0.40 **
3	0.93	(0.92 – 0.95)	0.98	0.70	0.06	0.46 **
5	0.79	(0.75 – 0.82)	0.97	0.50	0.16	0.32 **
10	0.65	(0.61 – 0.69)	0.89	0.40	0.18	0.26

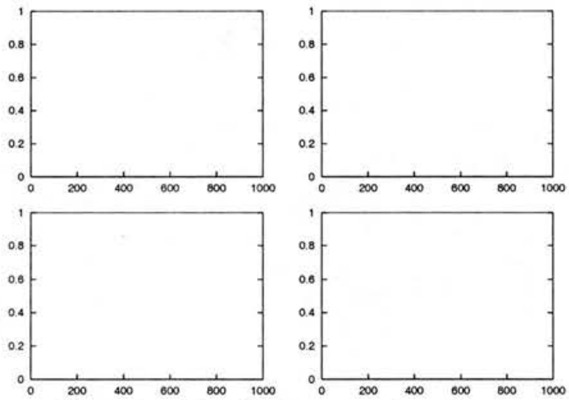
Summary of the final values of κ.

Figure C.14: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**proportional** ($p = 0.5$)
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



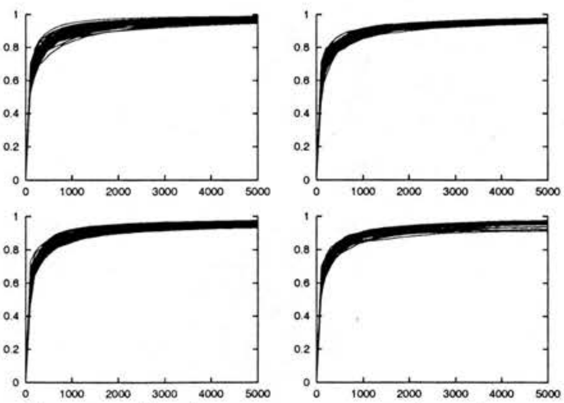
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.7)
2	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.98 **
3	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.98 **
5	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.98 **
10	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.98 **

Summary of the final values of σ.



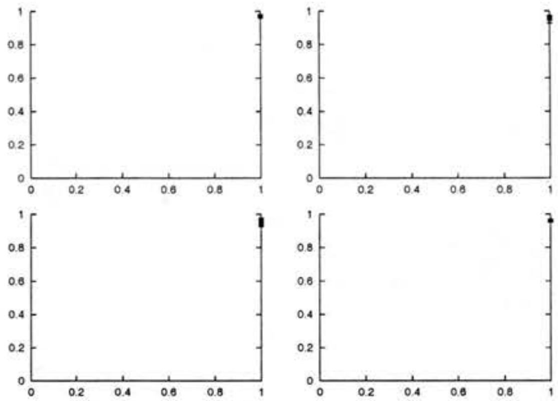
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.7)
2	0.97	(0.97 – 0.97)	0.99	0.95	0.01	0.46 **
3	0.97	(0.96 – 0.97)	0.98	0.94	0.01	0.72 **
5	0.96	(0.96 – 0.96)	0.98	0.93	0.01	0.94 **
10	0.96	(0.95 – 0.96)	0.98	0.91	0.02	0.98 **

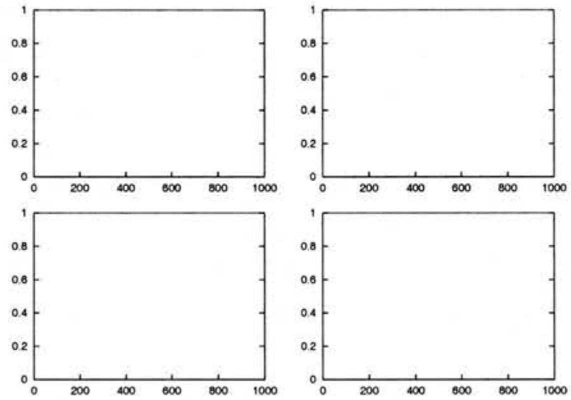
Summary of the final values of κ.

Figure C.15: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**identical random**
World:	**random**
Experiences:	**identical**
Hearer's Concept Creation driven by:	**discrimination**



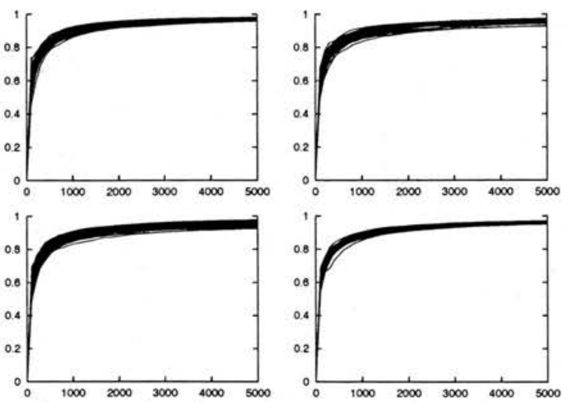
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.8)
2	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.94 **
3	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.98 **
5	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.98 **
10	1.00	(1.00 – 1.00)	1.00	1.00	0.00	0.98 **

Summary of the final values of σ.



Communicative success κ against time.

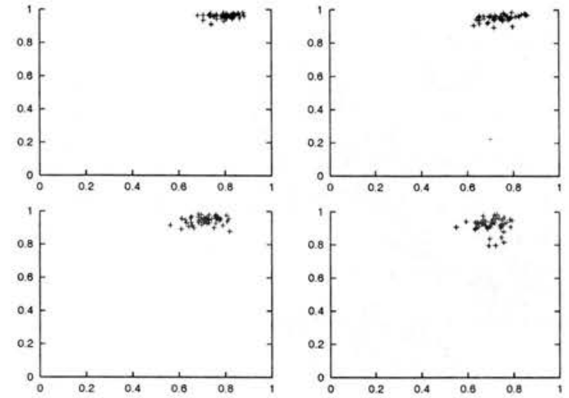
Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.8)
2	0.97	(0.97 – 0.97)	0.98	0.96	0.01	0.44 **
3	0.96	(0.96 – 0.97)	0.97	0.93	0.01	0.74 **
5	0.96	(0.96 – 0.97)	0.98	0.92	0.01	0.94 **
10	0.96	(0.96 – 0.96)	0.97	0.95	0.01	0.98 **

Summary of the final values of κ.

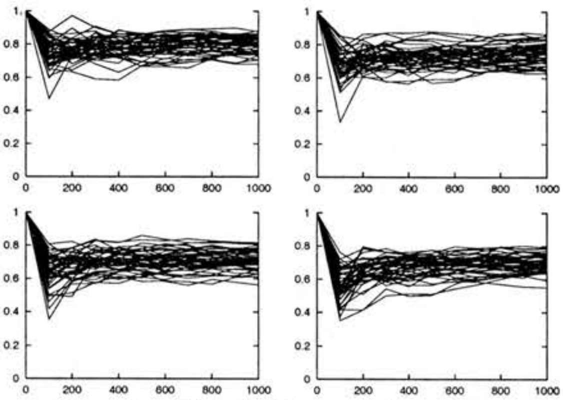
Figure C.16: Meaning similarity σ, communicative success κ (see box for parameters).

C.3 Clumpy World

Tree Growth Strategy:	**probabilistic**
Biases:	**uniform**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



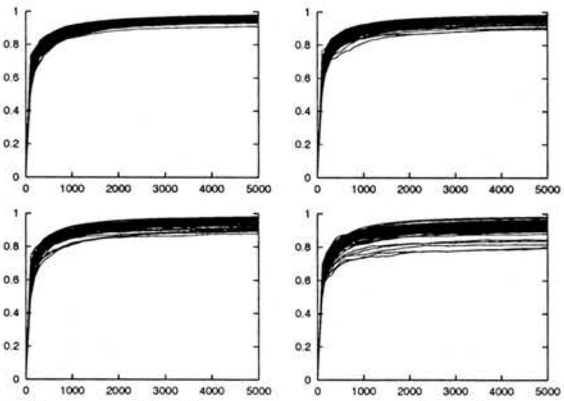
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.1)
2	0.80	(0.79 – 0.81)	0.88	0.68	0.06	0.14
3	0.75	(0.73 – 0.76)	0.86	0.63	0.08	0.10
5	0.71	(0.70 – 0.73)	0.82	0.56	0.08	0.20
10	0.70	(0.69 – 0.72)	0.79	0.55	0.07	0.48 **

Summary of the final values of σ.



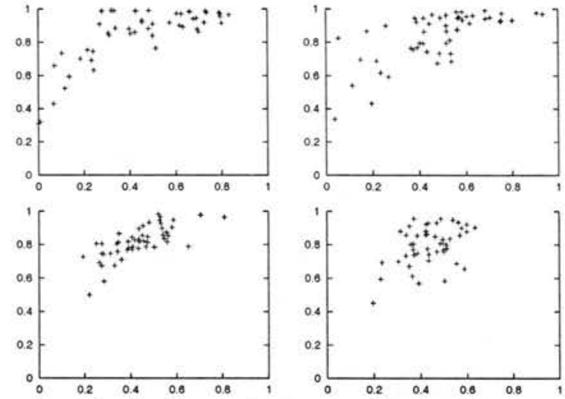
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.1)
2	0.96	(0.96 – 0.96)	0.98	0.91	0.01	0.42 **
3	0.95	(0.95 – 0.96)	0.98	0.89	0.02	0.44 **
5	0.94	(0.94 – 0.95)	0.98	0.88	0.03	0.64 **
10	0.92	(0.91 – 0.93)	0.98	0.80	0.04	0.88 **

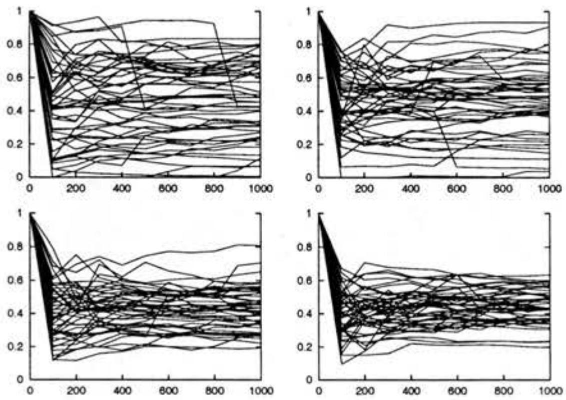
Summary of the final values of κ.

Figure C.17: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**random**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



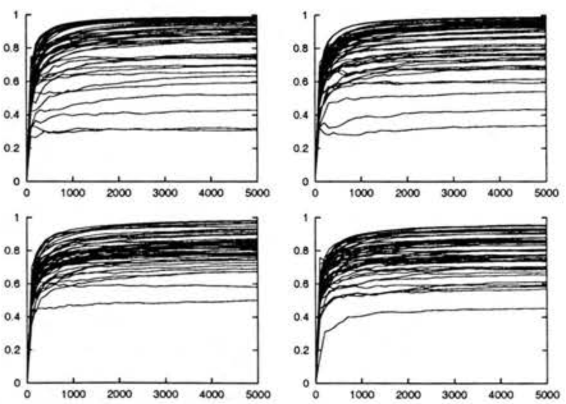
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.2)
2	0.45	(0.39 – 0.52)	0.83	0.00	0.52	0.14
3	0.48	(0.42 – 0.54)	0.93	0.04	0.42	0.20
5	0.44	(0.40 – 0.47)	0.81	0.19	0.29	0.14
10	0.44	(0.41 – 0.47)	0.63	0.20	0.22	0.26

Summary of the final values of σ.



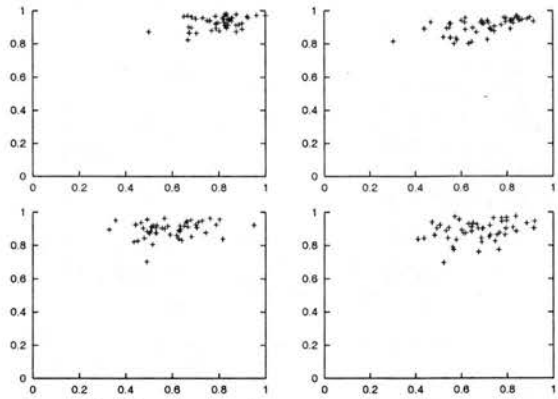
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.2)
2	0.85	(0.80 – 0.89)	0.99	0.31	0.20	0.28 *
3	0.84	(0.80 – 0.88)	0.99	0.34	0.18	0.32 **
5	0.82	(0.79 – 0.84)	0.98	0.50	0.12	0.42 **
10	0.80	(0.76 – 0.83)	0.96	0.45	0.14	0.58 **

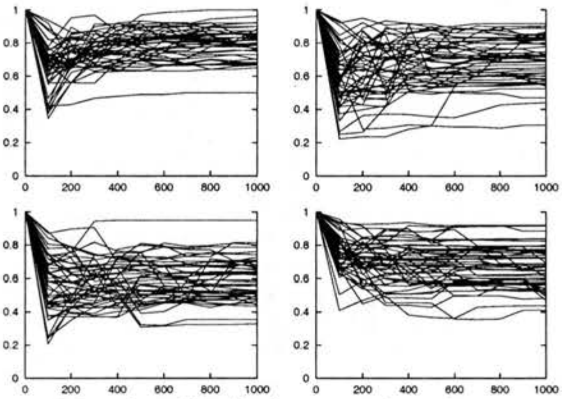
Summary of the final values of κ.

Figure C.18: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**proportional** ($p = 0.5$)
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



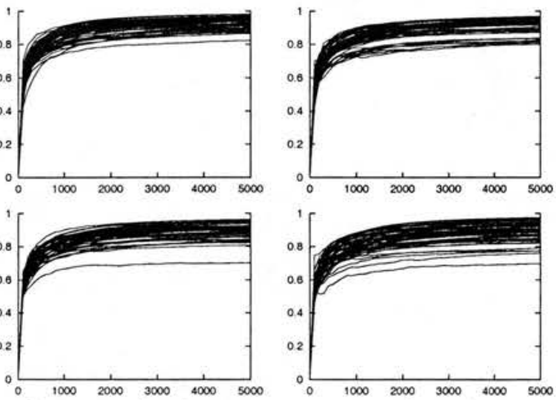
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.3)
2	0.80	(0.78 – 0.83)	1.00	0.50	0.11	0.20
3	0.70	(0.66 – 0.73)	0.92	0.30	0.19	0.38 **
5	0.60	(0.56 – 0.63)	0.95	0.33	0.20	0.60 **
10	0.68	(0.64 – 0.71)	0.92	0.41	0.19	0.24

Summary of the final values of σ.



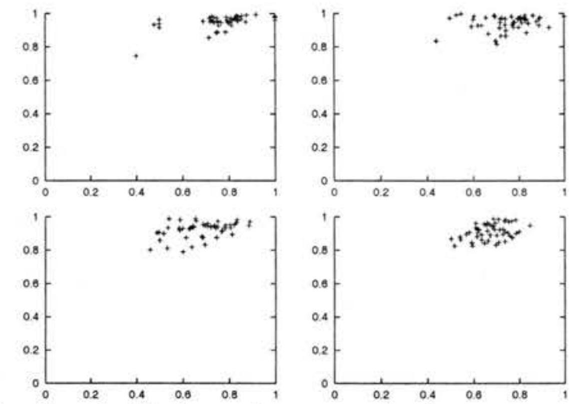
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.3)
2	0.93	(0.92 – 0.94)	0.98	0.82	0.04	0.40 **
3	0.91	(0.89 – 0.92)	0.97	0.80	0.05	0.36 **
5	0.90	(0.88 – 0.91)	0.96	0.70	0.05	0.12
10	0.89	(0.87 – 0.90)	0.98	0.70	0.07	0.38 **

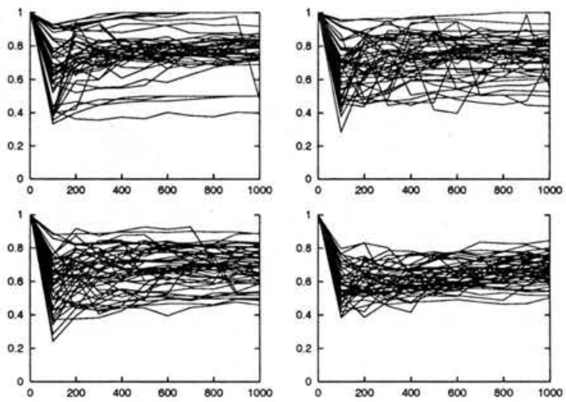
Summary of the final values of κ.

Figure C.19: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy: **probabilistic**
Biases: **identical random**
World: **clumpy**
Experiences: **different**
Hearer's Concept Creation
driven by: **discrimination**



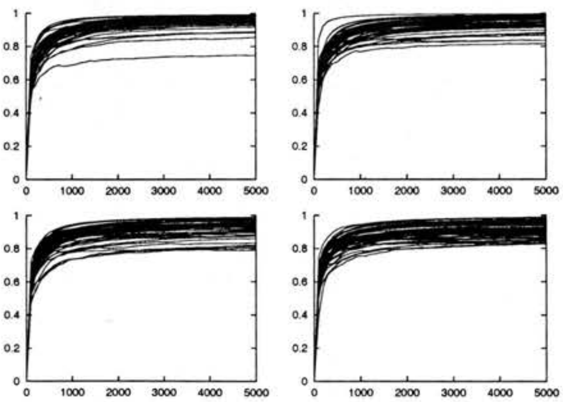
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.4)
2	0.78	(0.74 – 0.82)	1.00	0.40	0.17	0.26
3	0.75	(0.71 – 0.78)	1.00	0.44	0.15	0.16
5	0.68	(0.65 – 0.71)	0.89	0.46	0.16	0.18
10	0.67	(0.65 – 0.69)	0.85	0.50	0.11	0.30 *

Summary of the final values of σ.



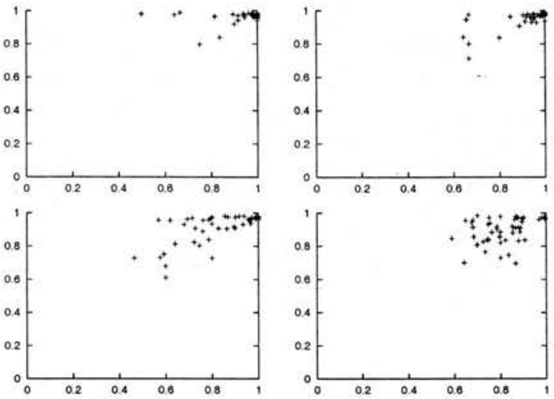
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.4)
2	0.95	(0.94 – 0.96)	0.99	0.75	0.04	0.18
3	0.94	(0.93 – 0.95)	1.00	0.82	0.05	0.38 **
5	0.92	(0.91 – 0.94)	0.99	0.79	0.05	0.42 **
10	0.91	(0.90 – 0.93)	0.99	0.83	0.05	0.52 **

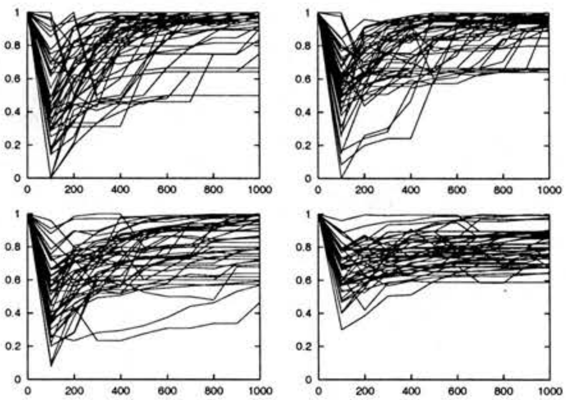
Summary of the final values of κ.

Figure C.20: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**uniform**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



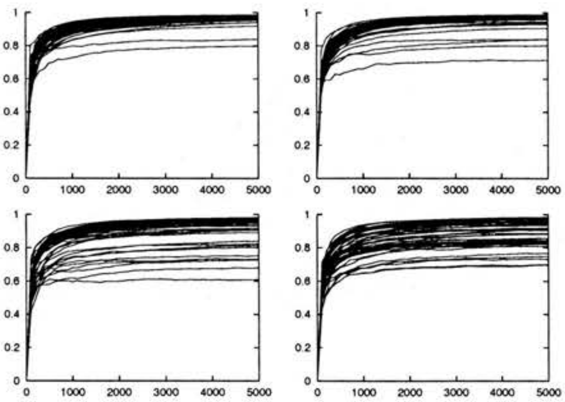
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.5)
2	0.94	(0.90 – 0.97)	1.00	0.50	0.13	0.54 **
3	0.93	(0.91 – 0.96)	1.00	0.64	0.11	0.80 **
5	0.83	(0.79 – 0.87)	1.00	0.46	0.17	0.68 **
10	0.81	(0.78 – 0.84)	1.00	0.59	0.12	0.96 **

Summary of the final values of σ.



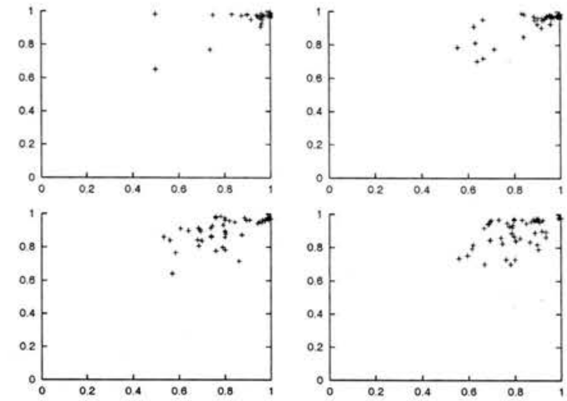
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.5)
2	0.97	(0.96 – 0.98)	0.99	0.80	0.03	0.38 **
3	0.95	(0.94 – 0.97)	0.99	0.71	0.05	0.72 **
5	0.91	(0.89 – 0.94)	0.98	0.61	0.10	0.70 **
10	0.89	(0.87 – 0.91)	0.98	0.70	0.09	0.92 **

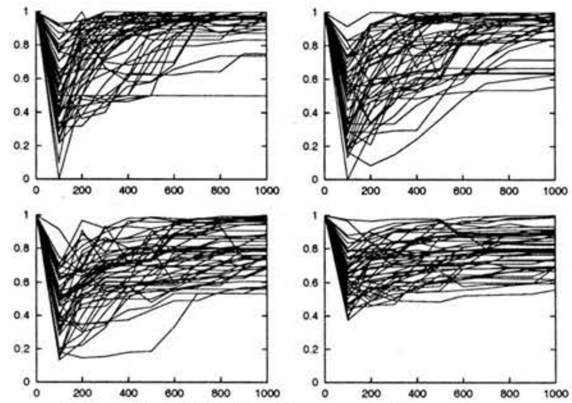
Summary of the final values of κ.

Figure C.21: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**random**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



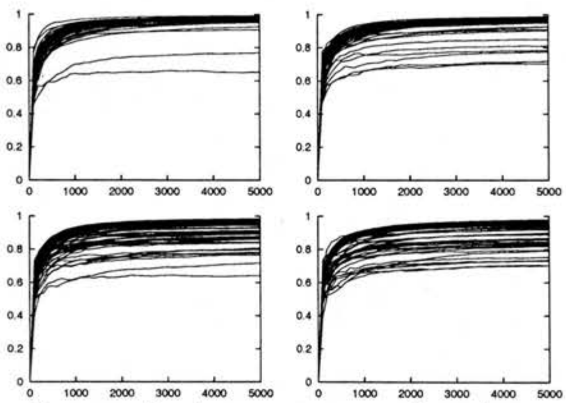
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.6)
2	0.95	(0.92 – 0.98)	1.00	0.50	0.11	0.56 **
3	0.91	(0.88 – 0.95)	1.00	0.56	0.13	0.72 **
5	0.80	(0.77 – 0.84)	1.00	0.53	0.16	0.82 **
10	0.81	(0.78 – 0.84)	1.00	0.56	0.14	0.91 **

Summary of the final values of σ.



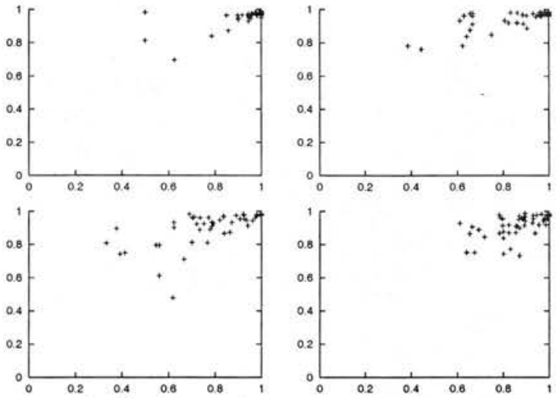
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.6)
2	0.96	(0.95 – 0.98)	0.99	0.65	0.06	0.50 **
3	0.94	(0.92 – 0.96)	0.99	0.70	0.07	0.58 **
5	0.91	(0.88 – 0.93)	0.98	0.64	0.08	0.64 **
10	0.89	(0.87 – 0.92)	0.98	0.70	0.09	0.83 **

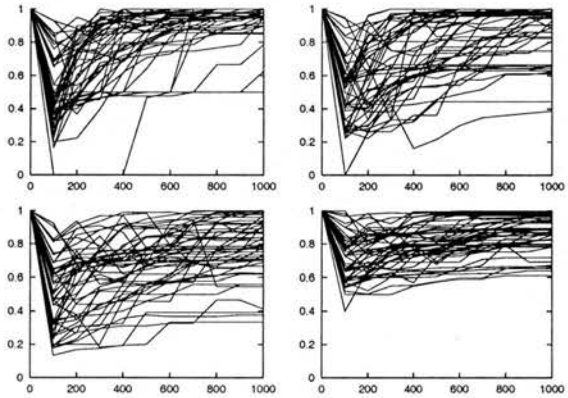
Summary of the final values of κ.

Figure C.22: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**proportional** ($p = 0.5$)
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



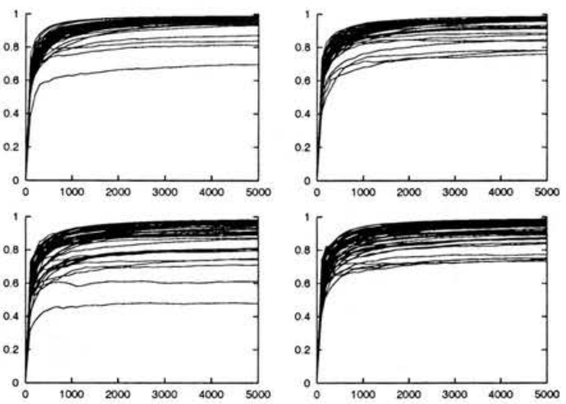
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.7)
2	0.94	(0.91 – 0.97)	1.00	0.50	0.12	0.58 **
3	0.87	(0.83 – 0.92)	1.00	0.39	0.18	0.56 **
5	0.78	(0.73 – 0.83)	1.00	0.33	0.22	0.68 **
10	0.86	(0.83 – 0.89)	1.00	0.61	0.12	0.90 **

Summary of the final values of σ.



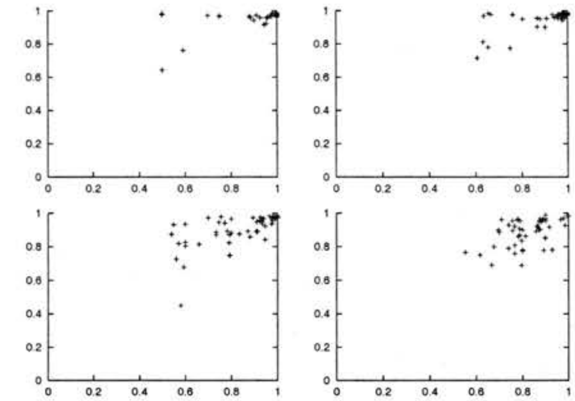
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.7)
2	0.96	(0.94 – 0.97)	0.99	0.70	0.05	0.30 *
3	0.94	(0.93 – 0.96)	0.99	0.76	0.06	0.50 **
5	0.90	(0.87 – 0.93)	0.98	0.48	0.11	0.68 **
10	0.91	(0.90 – 0.93)	0.99	0.73	0.07	0.84 **

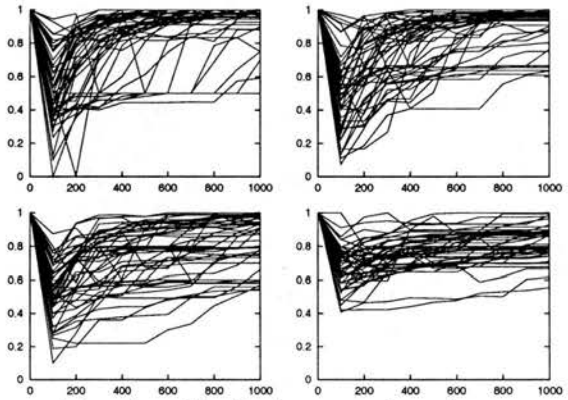
Summary of the final values of κ.

Figure C.23: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**identical random**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**discrimination**



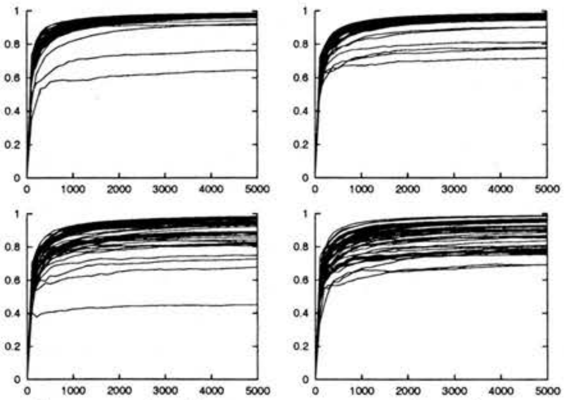
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.8)
2	0.93	(0.89 – 0.97)	1.00	0.50	0.15	0.56 **
3	0.92	(0.89 – 0.95)	1.00	0.61	0.13	0.76 **
5	0.83	(0.79 – 0.87)	1.00	0.54	0.18	0.72 **
10	0.82	(0.80 – 0.85)	1.00	0.55	0.12	0.92 **

Summary of the final values of σ.



Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.8)
2	0.96	(0.95 – 0.98)	0.99	0.64	0.06	0.42 **
3	0.95	(0.94 – 0.97)	0.99	0.71	0.06	0.70 **
5	0.90	(0.87 – 0.93)	0.98	0.45	0.11	0.62 **
10	0.89	(0.87 – 0.91)	0.99	0.69	0.09	0.68 **

Summary of the final values of κ.

Figure C.24: Meaning similarity σ, communicative success κ (see box for parameters).

APPENDIX D

Mutual Exclusivity Results

This appendix contains comprehensive results for the experiments reported in chapter 9, where levels of meaning similarity and communicative success are calculated under the following different parameters:

Tree Growth Strategy: the strategy used by the agents in creating meanings following discrimination failure;

Cognitive Biases: the method of bias allocation for the agents' biases;

Structure of the World: the method of constructing the agents' environment;

Agents' Experiences: whether the agents have the same or different interactions with their environment;

Hearer's Meaning Creation: in appendix D, the hearer's meaning creation is driven by interpretation failure and the assumption of mutual exclusivity in all experiments.

The layout of the results follows the same pattern as seen in appendix C; for each experiment, the following summary figures and tables are shown:

- at the top right, a scatter plot shows the relationship between meaning similarity σ and communicative success κ at the end of each simulation run;
- down the left, line plots show the progression of meaning similarity σ (upper), and of communicative success κ (lower) over time; each simulation run is shown with a separate line;

- to the right of these plots, the tables summarise the final results obtained, showing the average, range and variation in the values of σ and κ respectively.

Each figure itself consists of four sub-figures, which show results for experiments conducted with different numbers of sensory channels/features, as follows:

upper left objects are defined by *two* feature values, and agents have two corresponding sensory channels;

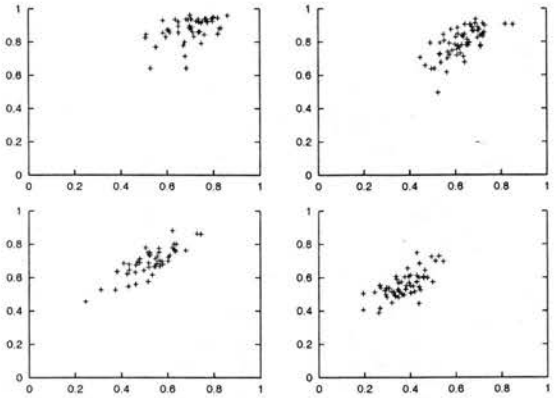
upper right objects are defined by *three* feature values, and agents have three corresponding sensory channels;

lower left objects are defined by *five* feature values, and agents have five corresponding sensory channels;

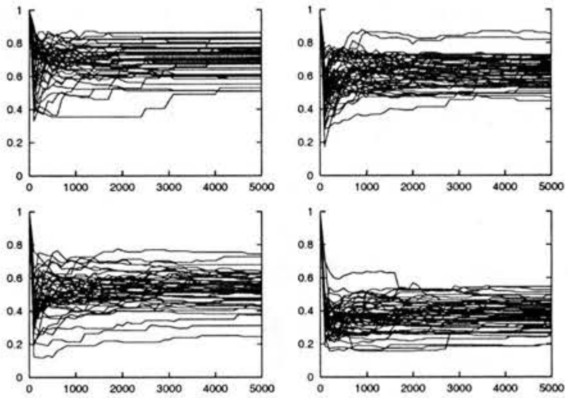
lower right objects are defined by *ten* feature values, and agents have ten corresponding sensory channels.

D.1 Mutual Exclusivity in a Random World

Tree Growth Strategy:	**probabilistic**
Biases:	**uniform**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



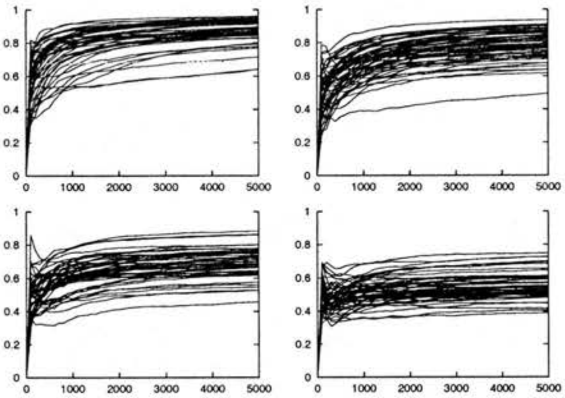
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.1)
2	0.70	(0.68 – 0.73)	0.86	0.51	0.12	0.52 **
3	0.63	(0.61 – 0.65)	0.85	0.45	0.13	0.61 **
5	0.53	(0.50 – 0.56)	0.74	0.25	0.18	0.82 **
10	0.38	(0.35 – 0.40)	0.55	0.19	0.21	0.92 **

Summary of the final values of σ.



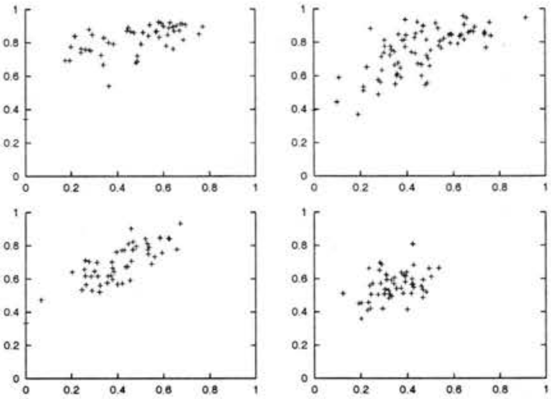
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.1)
2	0.88	(0.86 – 0.90)	0.96	0.64	0.08	0.68 **
3	0.80	(0.77 – 0.82)	0.94	0.49	0.11	0.84 **
5	0.70	(0.67 – 0.72)	0.88	0.46	0.12	0.94 **
10	0.56	(0.54 – 0.58)	0.75	0.39	0.14	0.95 **

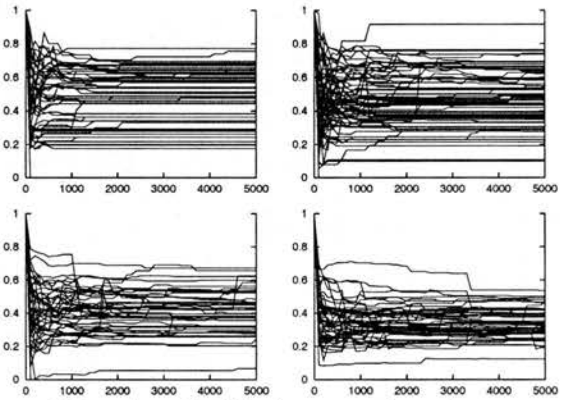
Summary of the final values of κ.

Figure D.1: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



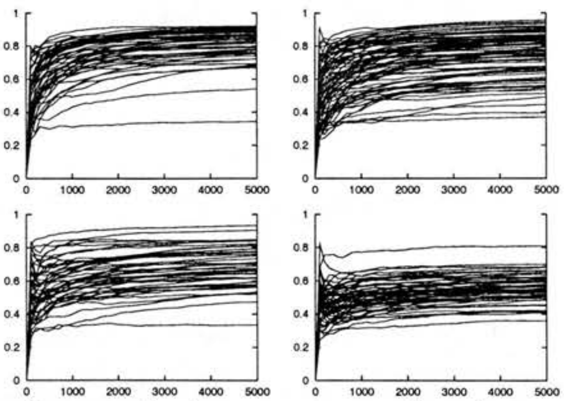
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.2)
2	0.47	(0.42 – 0.52)	0.77	0.00	0.38	0.24
3	0.46	(0.42 – 0.50)	0.92	0.00	0.38	0.15
5	0.41	(0.37 – 0.44)	0.67	0.00	0.34	0.10
10	0.34	(0.32 – 0.36)	0.53	0.12	0.26	0.26 *

Summary of the final values of σ.



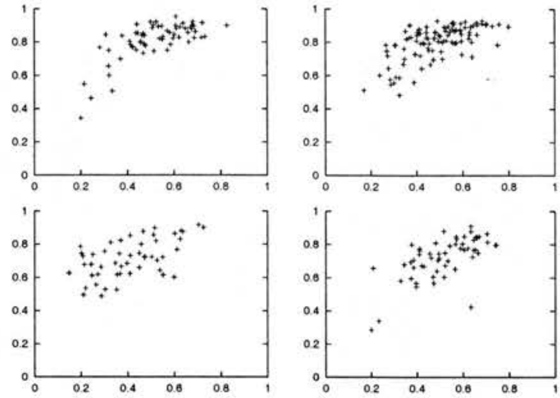
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.2)
2	0.81	(0.78 – 0.84)	0.92	0.34	0.13	0.36 **
3	0.75	(0.72 – 0.78)	0.96	0.37	0.18	0.21
5	0.69	(0.66 – 0.72)	0.93	0.33	0.17	0.16
10	0.56	(0.53 – 0.58)	0.81	0.36	0.15	0.39 **

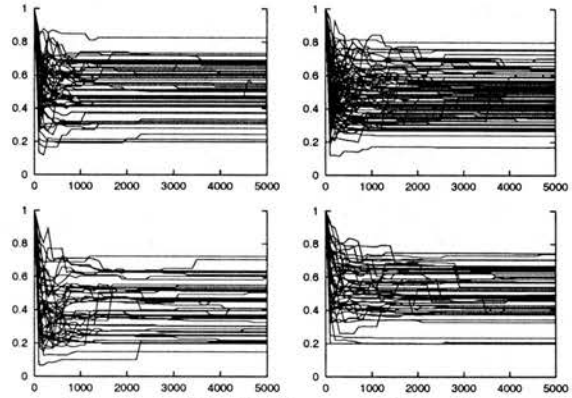
Summary of the final values of κ.

Figure D.2: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**proportional** ($p = 0.5$)
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



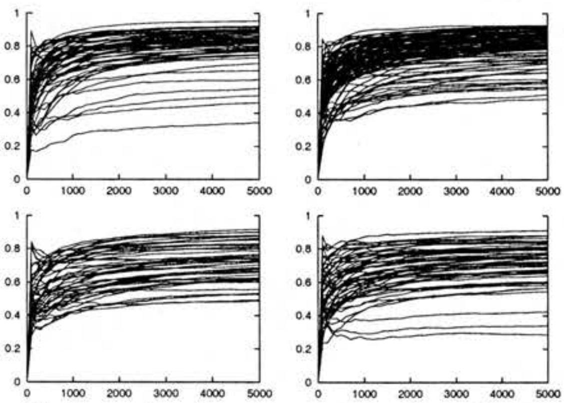
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.3)
2	0.52	(0.48 – 0.55)	0.83	0.20	0.27	0.90 **
3	0.49	(0.47 – 0.52)	0.80	0.17	0.27	0.90 **
5	0.41	(0.37 – 0.45)	0.72	0.15	0.36	0.90 **
10	0.52	(0.48 – 0.55)	0.74	0.20	0.25	0.64 **

Summary of the final values of σ.



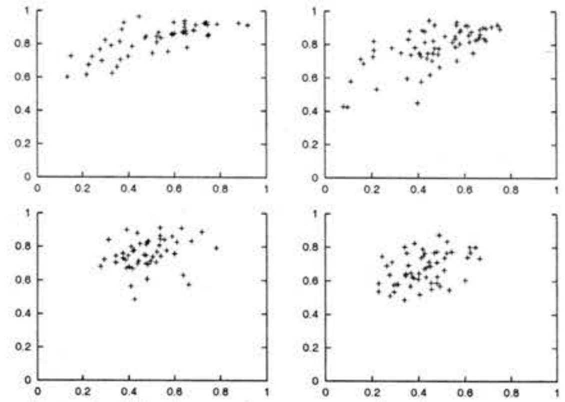
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.3)
2	0.81	(0.78 – 0.84)	0.95	0.34	0.14	0.96 **
3	0.80	(0.78 – 0.82)	0.93	0.48	0.13	0.83 **
5	0.72	(0.68 – 0.75)	0.92	0.49	0.16	0.74 **
10	0.72	(0.69 – 0.75)	0.91	0.29	0.17	0.61 **

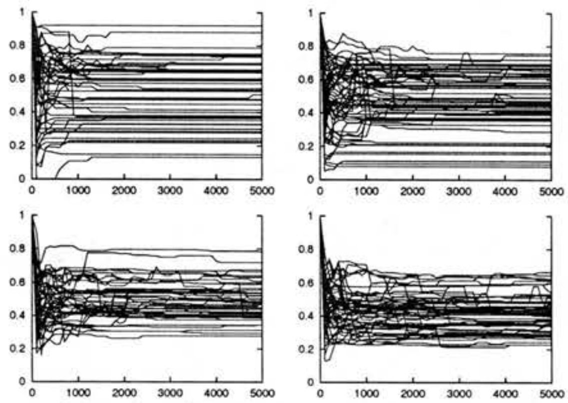
Summary of the final values of κ.

Figure D.3: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**identical random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



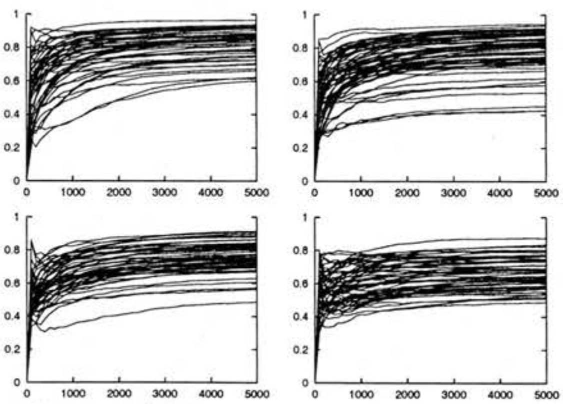
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.4)
2	0.53	(0.47 – 0.58)	0.92	0.13	0.36	0.64 **
3	0.48	(0.44 – 0.52)	0.76	0.08	0.35	0.58 **
5	0.48	(0.45 – 0.51)	0.78	0.28	0.22	0.58 **
10	0.42	(0.39 – 0.45)	0.66	0.23	0.25	0.69 **

Summary of the final values of σ.



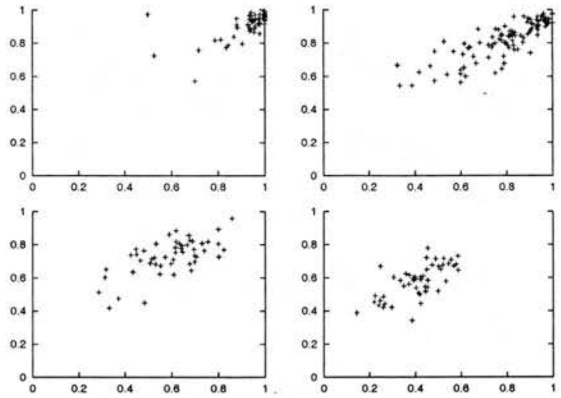
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.4)
2	0.83	(0.80 – 0.86)	0.97	0.60	0.11	0.76 **
3	0.79	(0.76 – 0.82)	0.94	0.42	0.16	0.61 **
5	0.76	(0.73 – 0.78)	0.91	0.49	0.12	0.68 **
10	0.67	(0.65 – 0.70)	0.87	0.49	0.14	0.78 **

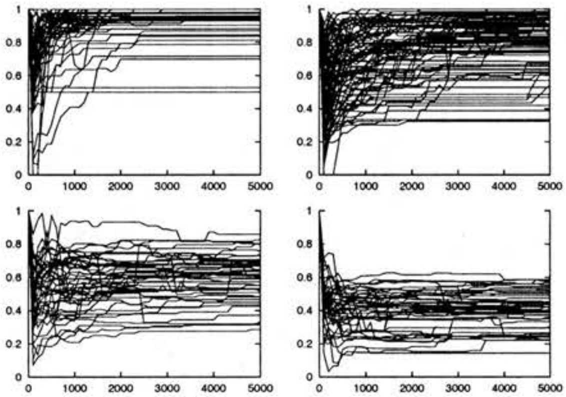
Summary of the final values of κ.

Figure D.4: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**uniform**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



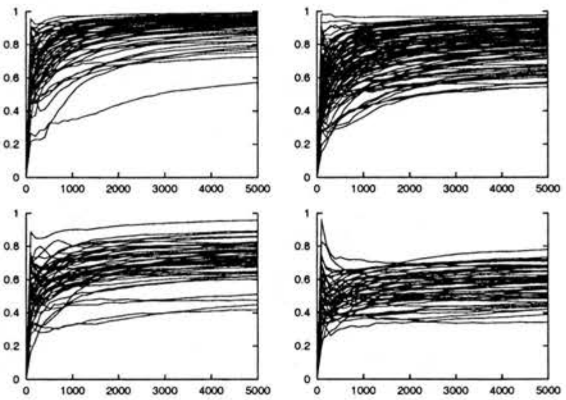
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.5)
2	0.93	(0.90 – 0.95)	1.00	0.50	0.11	0.36 **
3	0.80	(0.77 – 0.83)	1.00	0.32	0.20	0.48 **
5	0.59	(0.56 – 0.63)	0.86	0.29	0.23	0.40 **
10	0.41	(0.38 – 0.44)	0.59	0.14	0.26	0.34 **

Summary of the final values of σ.



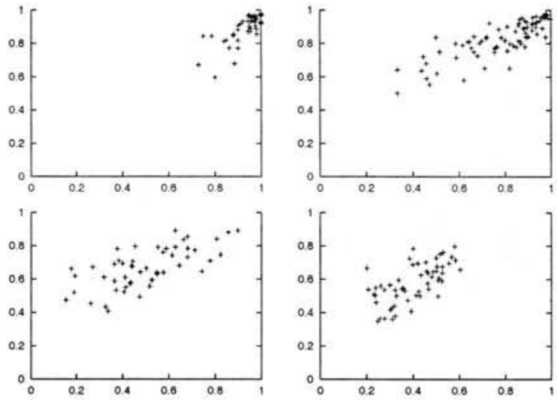
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.5)
2	0.91	(0.89 – 0.93)	0.99	0.57	0.08	0.35 **
3	0.82	(0.80 – 0.85)	0.98	0.54	0.14	0.21
5	0.73	(0.70 – 0.76)	0.96	0.42	0.15	0.24
10	0.58	(0.55 – 0.61)	0.78	0.34	0.17	0.14

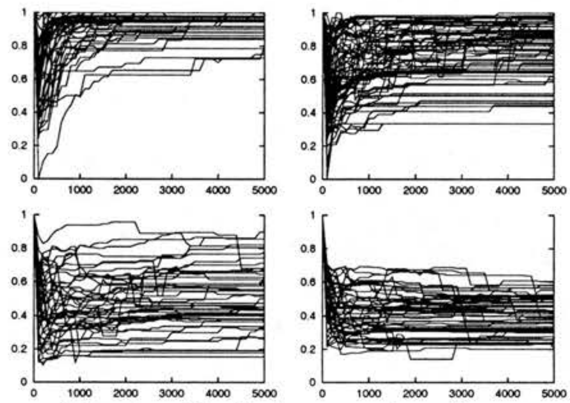
Summary of the final values of κ.

Figure D.5: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



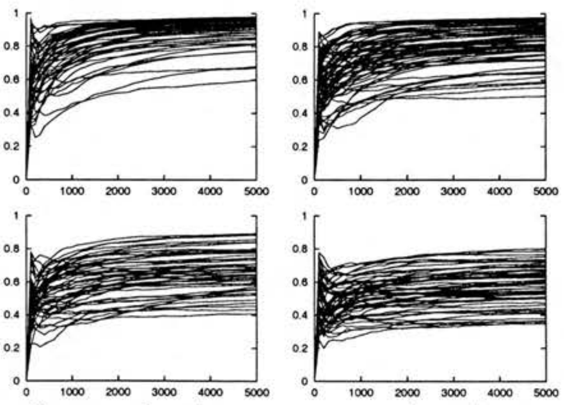
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.6)
2	0.94	(0.92 – 0.96)	1.00	0.73	0.07	0.38 **
3	0.78	(0.75 – 0.82)	1.00	0.33	0.22	0.42 **
5	0.50	(0.45 – 0.55)	0.90	0.15	0.36	0.20
10	0.40	(0.37 – 0.43)	0.60	0.20	0.27	0.15

Summary of the final values of σ.



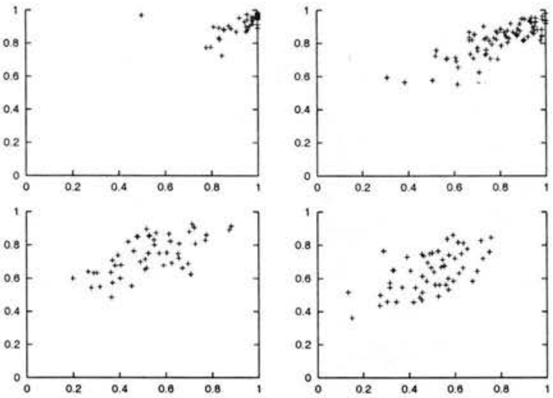
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.6)
2	0.90	(0.87 – 0.92)	0.98	0.60	0.09	0.32 **
3	0.83	(0.80 – 0.85)	0.98	0.50	0.13	0.16
5	0.67	(0.64 – 0.70)	0.89	0.41	0.18	0.20
10	0.58	(0.55 – 0.61)	0.80	0.35	0.20	0.18

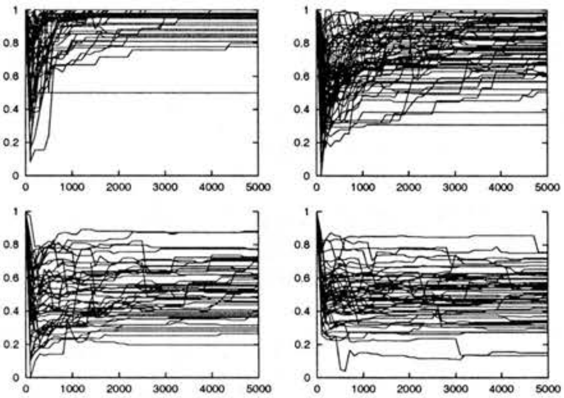
Summary of the final values of κ.

Figure D.6: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**proportional** ($p = 0.5$)
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



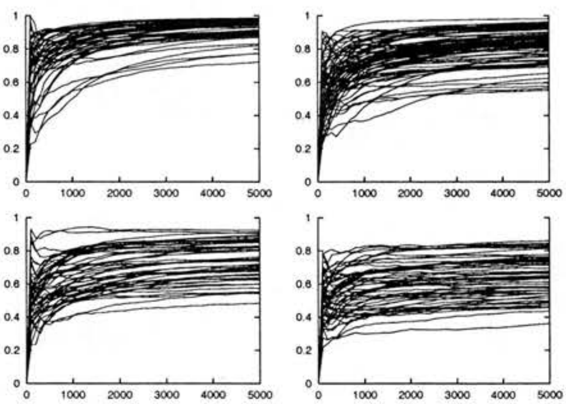
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.7)
2	0.94	(0.92 – 0.97)	1.00	0.50	0.10	0.52 **
3	0.80	(0.77 – 0.84)	1.00	0.31	0.19	0.46 **
5	0.53	(0.49 – 0.58)	0.88	0.20	0.30	0.26
10	0.50	(0.46 – 0.53)	0.76	0.13	0.28	0.15

Summary of the final values of σ.



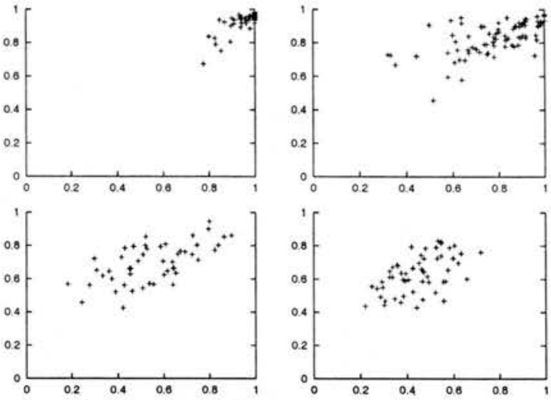
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.7)
2	0.92	(0.91 – 0.94)	0.99	0.72	0.06	0.26
3	0.82	(0.80 – 0.85)	0.98	0.55	0.12	0.31 **
5	0.74	(0.71 – 0.77)	0.93	0.48	0.15	0.12
10	0.64	(0.61 – 0.67)	0.86	0.36	0.19	0.21

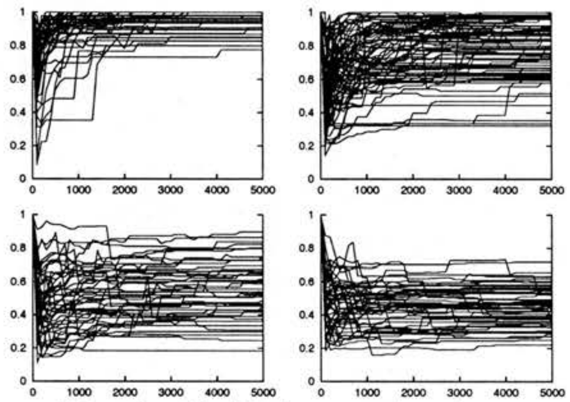
Summary of the final values of κ.

Figure D.7: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**identical random**
World:	**random**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



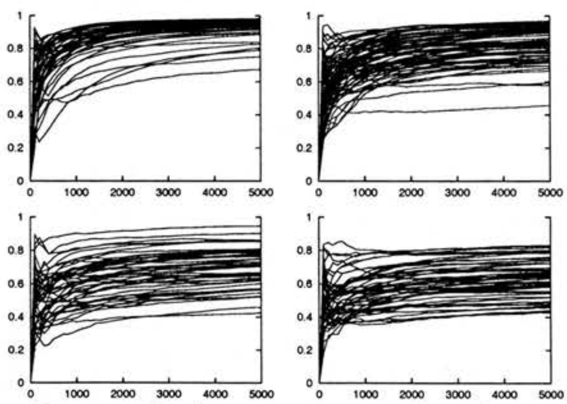
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.8)
2	0.95	(0.93 – 0.97)	1.00	0.78	0.06	0.48 **
3	0.79	(0.75 – 0.82)	1.00	0.32	0.21	0.47 **
5	0.55	(0.50 – 0.60)	0.90	0.18	0.31	0.26
10	0.45	(0.42 – 0.48)	0.72	0.22	0.25	0.11

Summary of the final values of σ.



Communicative success κ against time.

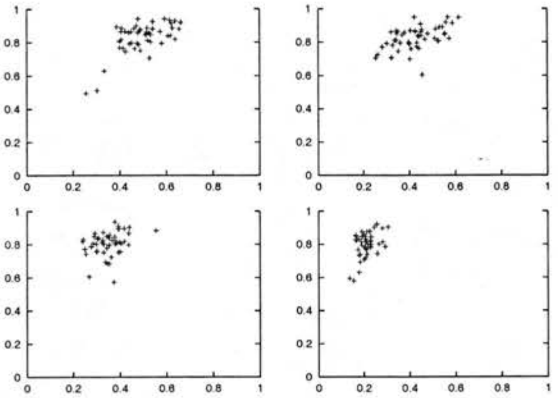
Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.8)
2	0.93	(0.91 – 0.94)	0.98	0.67	0.07	0.30 *
3	0.83	(0.81 – 0.85)	0.97	0.46	0.12	0.21
5	0.69	(0.66 – 0.73)	0.95	0.42	0.17	0.28 *
10	0.63	(0.60 – 0.66)	0.83	0.43	0.18	0.25

Summary of the final values of κ.

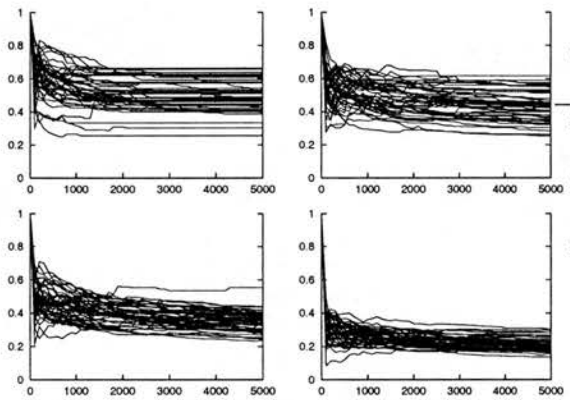
Figure D.8: Meaning similarity σ, communicative success κ (see box for parameters).

D.2 Mutual Exclusivity in a Clumpy World

Tree Growth Strategy:	**probabilistic**
Biases:	**uniform**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



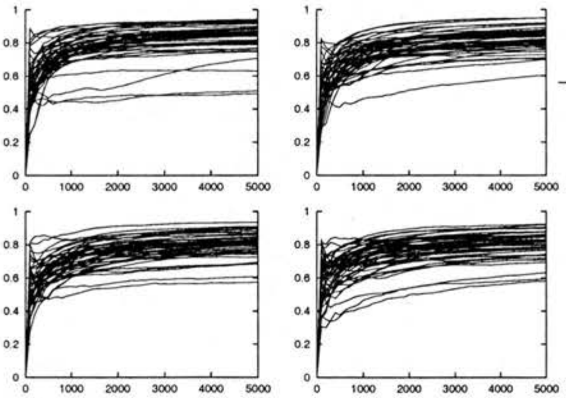
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.17)	KS (D.1)
2	0.50	(0.47 – 0.52)	0.66	0.26	0.18	1.0 **	0.74 **
3	0.43	(0.41 – 0.46)	0.62	0.25	0.21	1.0 **	0.73 **
5	0.35	(0.33 – 0.37)	0.55	0.24	0.17	1.0 **	0.82 **
10	0.21	(0.20 – 0.22)	0.31	0.14	0.17	1.0 **	0.85 **

Summary of the final values of σ.



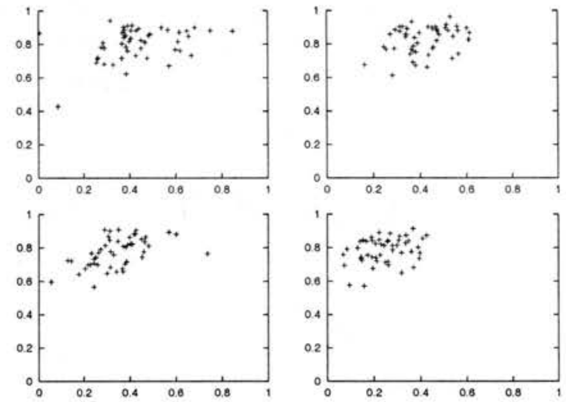
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.17)	KS (D.1)
2	0.83	(0.80 – 0.85)	0.94	0.49	0.11	0.94 **	0.32 **
3	0.83	(0.81 – 0.85)	0.95	0.60	0.08	0.88 **	0.22
5	0.81	(0.79 – 0.83)	0.93	0.57	0.09	0.90 **	0.62 **
10	0.80	(0.78 – 0.82)	0.92	0.58	0.09	0.80 **	0.87 **

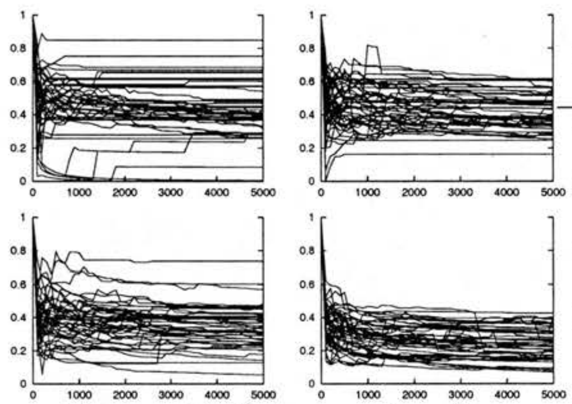
Summary of the final values of κ.

Figure D.9: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**random**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



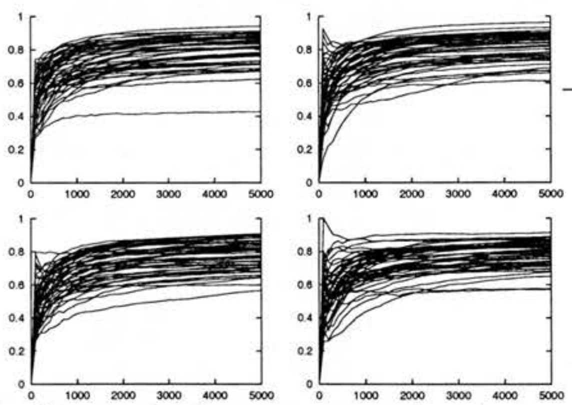
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.18)	KS (D.2)
2	0.43	(0.39 – 0.48)	0.85	0.01	0.36	0.20	0.26
3	0.42	(0.39 – 0.45)	0.62	0.16	0.25	0.28 *	0.23
5	0.34	(0.31 – 0.38)	0.74	0.05	0.35	0.36 **	0.28 *
10	0.24	(0.22 – 0.27)	0.43	0.07	0.39	0.72 **	0.47 **

Summary of the final values of σ.



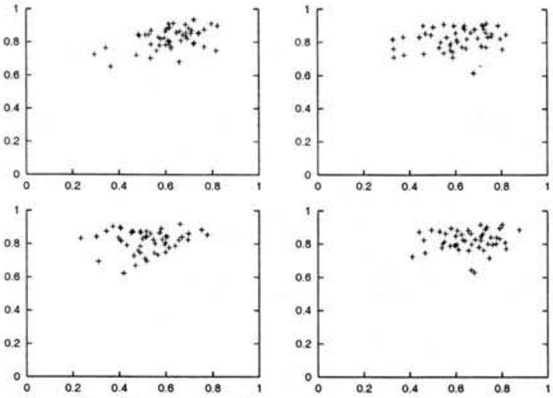
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.18)	KS (D.2)
2	0.80	(0.78 – 0.83)	0.94	0.43	0.12	0.44 **	0.12
3	0.82	(0.80 – 0.85)	0.96	0.61	0.10	0.42 **	0.28 *
5	0.78	(0.75 – 0.80)	0.91	0.57	0.11	0.24	0.36 **
10	0.78	(0.76 – 0.80)	0.91	0.57	0.10	0.26	0.87 **

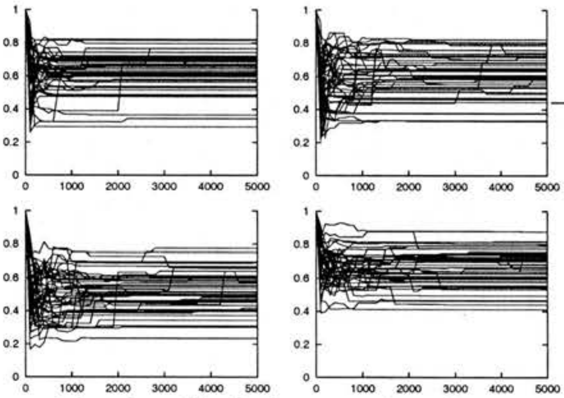
Summary of the final values of κ.

Figure D.10: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**proportional** ($p = 0.5$)
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



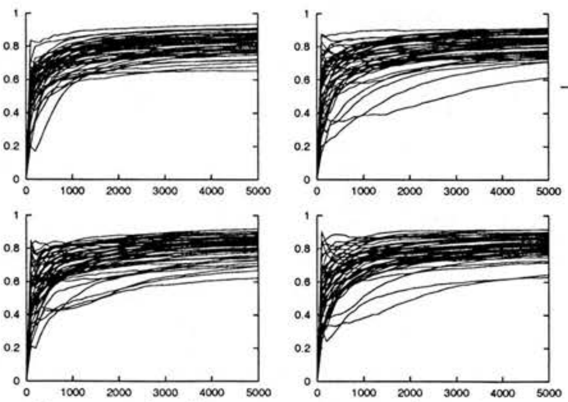
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.19)	KS (D.3)
2	0.62	(0.59 – 0.65)	0.82	0.29	0.18	0.68 **	0.37 **
3	0.61	(0.57 – 0.64)	0.82	0.33	0.21	0.30 *	0.41 **
5	0.52	(0.49 – 0.55)	0.78	0.23	0.22	0.32 **	0.40 **
10	0.66	(0.63 – 0.69)	0.88	0.41	0.16	0.14	0.45 **

Summary of the final values of σ.



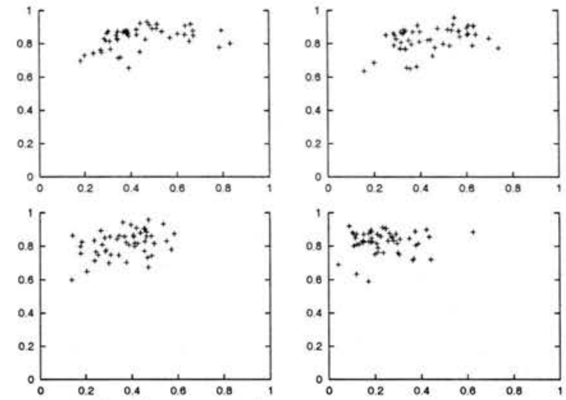
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.19)	KS (D.3)
2	0.82	(0.80 – 0.84)	0.94	0.65	0.08	0.74 **	0.16
3	0.82	(0.80 – 0.84)	0.91	0.61	0.08	0.60 **	0.14
5	0.82	(0.80 – 0.83)	0.92	0.62	0.08	0.62 **	0.50 **
10	0.82	(0.80 – 0.84)	0.92	0.63	0.07	0.48 **	0.46 **

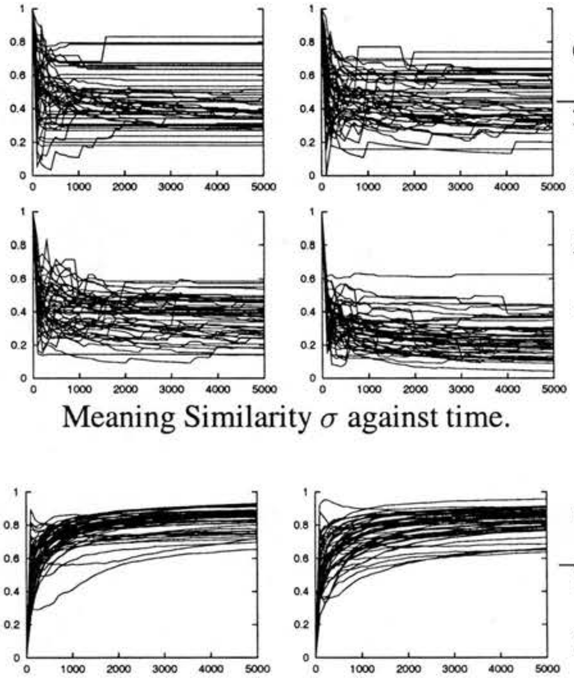
Summary of the final values of κ.

Figure D.11: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**probabilistic**
Biases:	**identical random**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



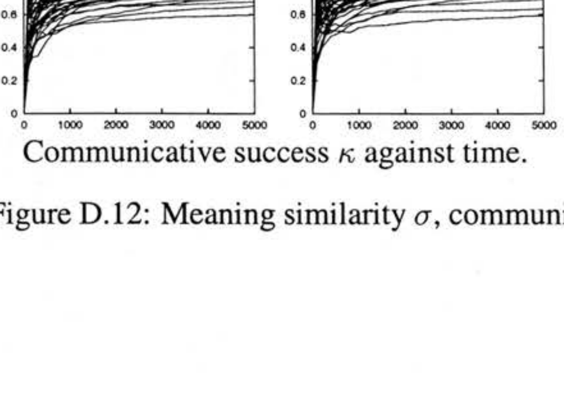
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.20)	KS (D.4)
2	0.44	(0.40 – 0.48)	0.83	0.18	0.35	0.84 **	0.34 **
3	0.44	(0.40 – 0.48)	0.74	0.16	0.32	0.78 **	0.26 *
5	0.37	(0.34 – 0.40)	0.59	0.14	0.31	0.86 **	0.40 **
10	0.24	(0.21 – 0.27)	0.63	0.04	0.47	0.98 **	0.62 **

Summary of the final values of σ.



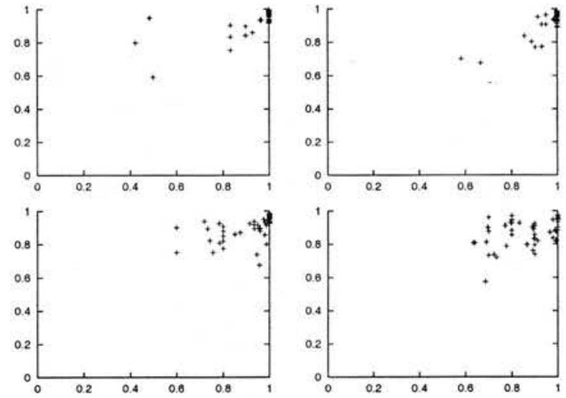
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.20)	KS (D.4)
2	0.84	(0.82 – 0.85)	0.93	0.65	0.08	0.86 **	0.18
3	0.82	(0.80 – 0.84)	0.96	0.64	0.09	0.80 **	0.25 *
5	0.82	(0.79 – 0.84)	0.96	0.60	0.09	0.68 **	0.30 *
10	0.82	(0.80 – 0.84)	0.92	0.59	0.08	0.60 **	0.69 **

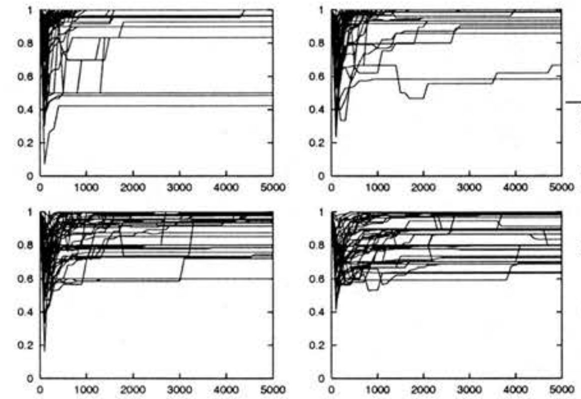
Summary of the final values of κ.

Figure D.12: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**uniform**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



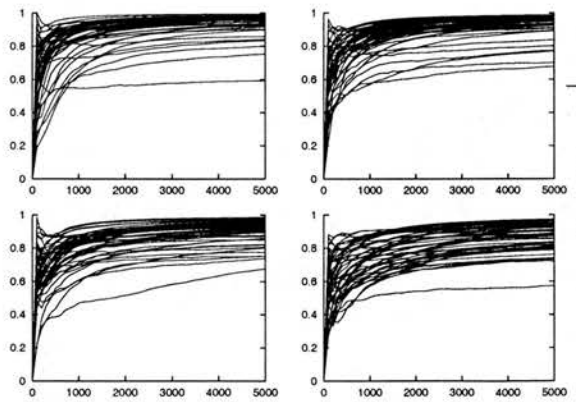
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.21)	KS (D.5)
2	0.95	(0.91 – 0.99)	1.00	0.42	0.14	0.28 *	0.48 **
3	0.97	(0.95 – 0.99)	1.00	0.58	0.08	0.38 **	0.75 **
5	0.92	(0.88 – 0.95)	1.00	0.60	0.12	0.32 **	0.80 **
10	0.85	(0.82 – 0.89)	1.00	0.63	0.13	0.32 **	0.98 **

Summary of the final values of σ.



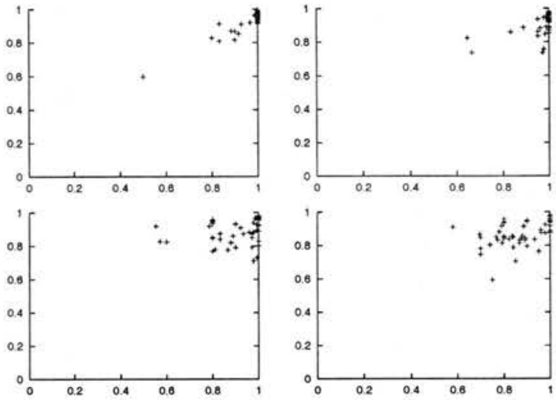
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.21)	KS (D.5)
2	0.94	(0.92 – 0.96)	0.99	0.59	0.08	0.26	0.34 **
3	0.93	(0.91 – 0.95)	0.99	0.68	0.07	0.36 **	0.56 **
5	0.90	(0.88 – 0.92)	0.99	0.67	0.08	0.36 **	0.72 **
10	0.86	(0.84 – 0.88)	0.97	0.57	0.10	0.24	0.94 **

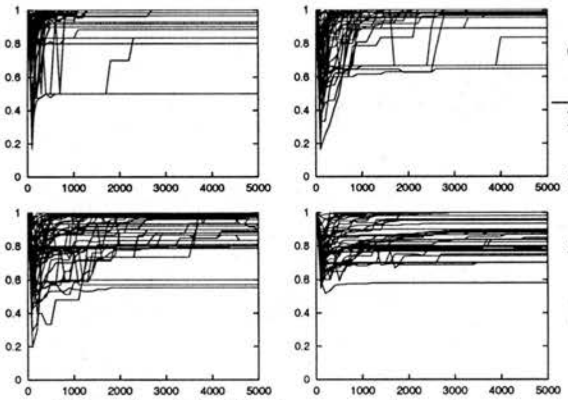
Summary of the final values of κ.

Figure D.13: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**random**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



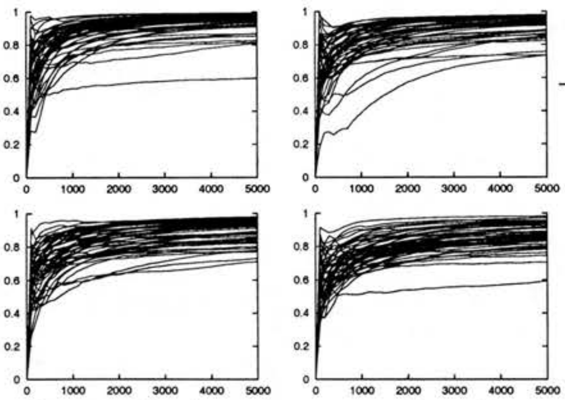
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.22)	KS (D.6)
2	0.97	(0.95 – 0.99)	1.00	0.50	0.09	0.22	0.56 **
3	0.97	(0.95 – 0.99)	1.00	0.65	0.07	0.42 **	0.79 **
5	0.91	(0.87 – 0.94)	1.00	0.55	0.13	0.38 **	0.86 **
10	0.87	(0.84 – 0.90)	1.00	0.58	0.12	0.28 *	0.96 **

Summary of the final values of σ.



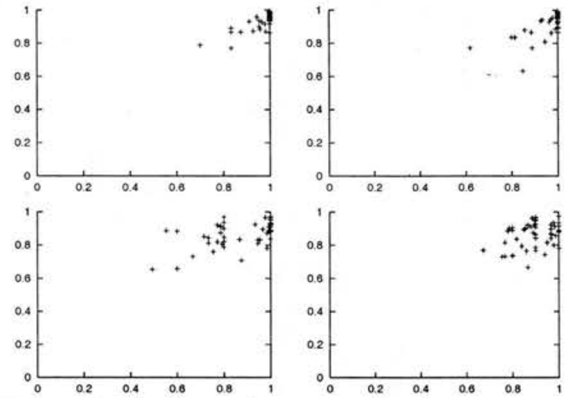
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.22)	KS (D.6)
2	0.94	(0.92 – 0.96)	0.99	0.60	0.07	0.42 **	0.36 **
3	0.92	(0.90 – 0.93)	0.98	0.74	0.07	0.46 **	0.47 **
5	0.89	(0.87 – 0.91)	0.98	0.71	0.08	0.18	0.74 **
10	0.87	(0.85 – 0.89)	0.98	0.59	0.08	0.30 *	0.90 **

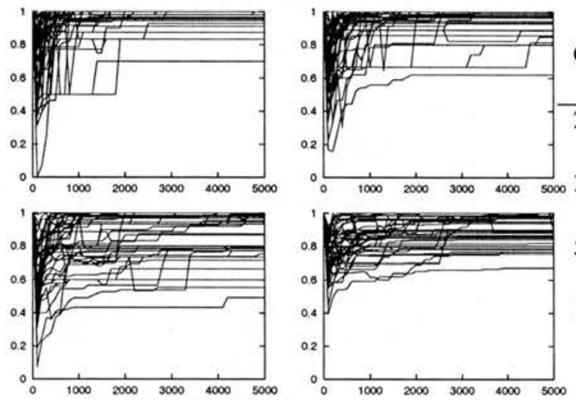
Summary of the final values of κ.

Figure D.14: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**proportional** ($p = 0.5$)
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



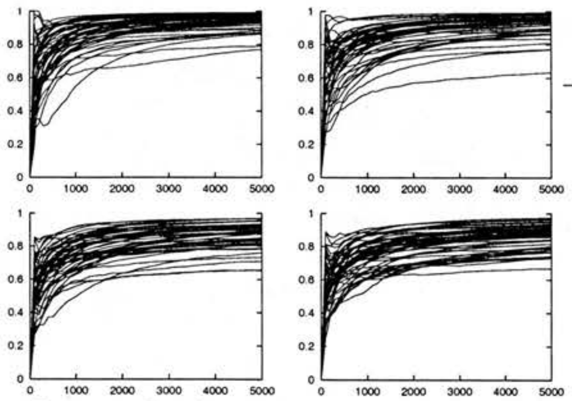
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.23)	KS (D.7)
2	0.97	(0.96 – 0.99)	1.00	0.70	0.06	0.28 *	0.26
3	0.97	(0.95 – 0.99)	1.00	0.62	0.07	0.38 **	0.67 **
5	0.88	(0.84 – 0.91)	1.00	0.49	0.16	0.34 **	0.80 **
10	0.90	(0.87 – 0.92)	1.00	0.67	0.09	0.22	0.96 **

Summary of the final values of σ.



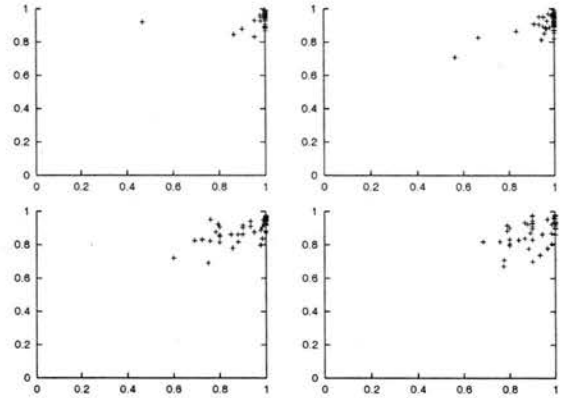
Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.23)	KS (D.7)
2	0.94	(0.92 – 0.95)	0.99	0.77	0.05	0.40 **	0.18
3	0.92	(0.90 – 0.94)	0.99	0.63	0.07	0.40 **	0.53 **
5	0.87	(0.85 – 0.89)	0.97	0.65	0.09	0.34 **	0.50 **
10	0.86	(0.84 – 0.88)	0.98	0.67	0.09	0.38 **	0.75 **

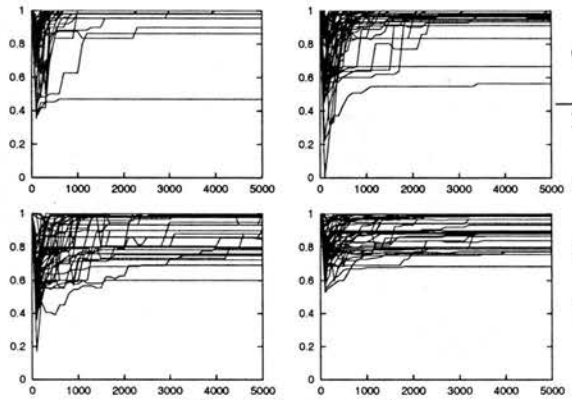
Summary of the final values of κ.

Figure D.15: Meaning similarity σ, communicative success κ (see box for parameters).

Tree Growth Strategy:	**intelligent**
Biases:	**identical random**
World:	**clumpy**
Experiences:	**different**
Hearer's Concept Creation driven by:	**communication**



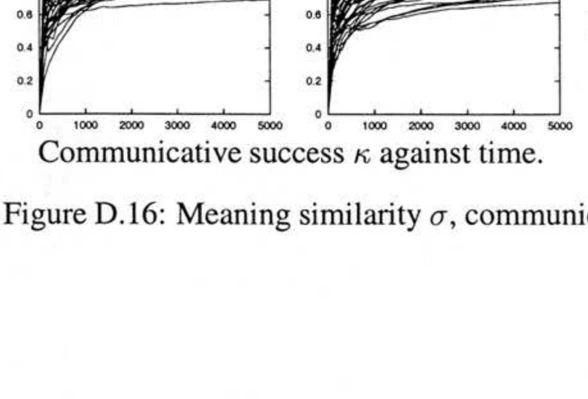
Meaning similarity and communicative success.



Meaning Similarity σ against time.

Ch	$\bar{\sigma}$	CI	Max (σ)	Min (σ)	CoV (σ)	KS (C.24)	KS (D.8)
2	0.98	(0.96 – 1.00)	1.00	0.47	0.08	0.30 *	0.48 **
3	0.97	(0.95 – 0.99)	1.00	0.56	0.08	0.44 **	0.74 **
5	0.92	(0.89 – 0.95)	1.00	0.60	0.11	0.40 **	0.80 **
10	0.92	(0.89 – 0.94)	1.00	0.68	0.09	0.46 **	0.98 **

Summary of the final values of σ.



Communicative success κ against time.

Ch	$\bar{\kappa}$	CI	Max (κ)	Min (κ)	CoV (κ)	KS (C.24)	KS (D.8)
2	0.95	(0.94 – 0.96)	0.99	0.83	0.04	0.34 **	0.34 **
3	0.92	(0.90 – 0.93)	0.99	0.71	0.06	0.56 **	0.47 **
5	0.90	(0.88 – 0.92)	0.98	0.69	0.07	0.24	0.80 **
10	0.88	(0.86 – 0.91)	0.98	0.67	0.09	0.16	0.81 **

Summary of the final values of κ.

Figure D.16: Meaning similarity σ, communicative success κ (see box for parameters).

Appendix E

Examples of Agents' Lexicons

This appendix contains examples of the lexicons which are created by the agents. As we saw in section 6.5, an agent's lexicon stores its interaction history in terms of signal-meaning associations. Each entry in the lexicon contains the following components:

- a signal s;
- a meaning m;
- a count of how many times the pair has been used u;
- a confidence probability p, which represents the agent's confidence in the association between the signal and meaning.

The size of a complete lexicon is potentially very large, containing every signal-meaning pair for which the agent has created an association. Table E.1, which extends over a number of pages, shows details of an agent's lexicon after 5000 communicative episodes. The agent has five sensory channels, but is using the intelligent tree growth strategy, and has not developed any conceptual structure on sensory channel 3. It is immediately apparent not only that the lexicon is very large, but also that the level of confidence in many of the signal-meaning pairs is very low. These particular associations are never likely to be used in the communication process, but are nevertheless maintained, because the lexicon contains a *complete* history of all the agent's communicative interactions.

Note also the entry for the association between the signal *klklv* and the meaning $4 - 1$; although the agent has only made this association once, there were no other competing semantic hypotheses in the episode when the signal was received, and so the agent is certain of the association. Associations like this, which have a high confidence probability p but a low usage u, however, are very unstable, as just one further exposure to the signal in another context, where the associated meaning (here $4-1$) is not a possible semantic hypothesis, would completely undermine the mapping.

In section 6.5.1, however, I described the introspective obverter algorithm, and introduced the notion of an *active lexicon*, or the lexical entries which, at any particular time, represent signal-meaning pairs which could be used by the agent when it chooses a signal. Introspective obverter is based on choosing a signal which the agent itself would interpret correctly, and so, at any time, there is only one meaning which the agent will use to interpret a particular signal[1]. Therefore, to derive the active lexicon from a complete lexicon, we perform the following process iteratively, until there are no entries left in the original lexicon:

1. Find the signal-meaning pair with the highest confidence probability.
2. Move this entry to the active lexicon.
3. Delete all entries in the original lexicon which contain the signal from the chosen pair.

Table E.2 shows the active lexicon which is derived from the complete lexicon in table E.1. There are of course fewer entries, and those which are included are those in which the agent has relatively high confidence, and so would prefer to use to interpret a signal.

[1] The actual interpretation will depend on the context in which the signal is presented, but this will be ignored here.

Table E.1: Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
ouy	0-0	118	0.323288
us	0-0	26	0.0641975
dt	0-0	1	0.0588235
qlokt	0-0	6	0.0588235
qj	0-0	14	0.0510949
ty	0-0	14	0.05
le	0-0	6	0.0454545
nvvb	0-0	12	0.0449438
ql	0-0	43	0.0388087
ms	0-0	10	0.0308642
kx	0-00	254	0.298824
isuyo	0-00	23	0.132948
wtz	0-00	46	0.126722
lho	0-00	85	0.111402
ms	0-00	36	0.111111
mtsb	0-00	36	0.108108
mtn	0-00	156	0.0977444
tg	0-00	41	0.0933941
ve	0-00	11	0.0808824
dt	0-00	1	0.0588235
co	0-000	57	0.311475
hp	0-000	18	0.163636
po	0-000	19	0.0678571
mtsb	0-000	18	0.0540541
cq	0-000	13	0.0439189
tg	0-000	19	0.0432802
jgy	0-000	27	0.0419907
ql	0-000	46	0.0415162
xc	0-000	40	0.0414079
ty	0-000	10	0.0357143
sz	0-0000	59	0.265766
cvf	0-0000	6	0.101695
cq	0-0000	10	0.0337838

continued on next page ...

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
lho	0-0000	25	0.0327654
ve	0-0000	4	0.0294118
xj	0-0000	5	0.027933
hp	0-0000	3	0.0272727
zg	0-0000	4	0.0272109
xq	0-0000	17	0.0271565
le	0-0000	2	0.0151515
cvf	0-0001	16	0.271186
sz	0-0001	45	0.202703
ylf	0-0001	16	0.063745
lc	0-0001	9	0.0436893
xj	0-0001	5	0.027933
zg	0-0001	4	0.0272109
mtsb	0-0001	8	0.024024
cq	0-0001	7	0.0236486
hp	0-0001	2	0.0181818
le	0-0001	2	0.0151515
co	0-001	44	0.240437
hp	0-001	22	0.2
cvf	0-001	10	0.169492
djh	0-001	19	0.134752
po	0-001	24	0.0857143
ve	0-001	9	0.0661765
ty	0-001	16	0.0571429
jgy	0-001	35	0.0544323
bv	0-001	23	0.0490405
tg	0-001	20	0.0455581
jgy	0-01	225	0.349922
kx	0-01	104	0.122353
cvf	0-01	7	0.118644
xq	0-01	70	0.111821
co	0-01	20	0.10929
bv	0-01	49	0.104478
dka	0-01	20	0.104167

continued on next page ...

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
px	0-01	61	0.102007
lc	0-01	20	0.0970874
ol	0-01	20	0.0947867
ba	0-010	112	0.34891
pgk	0-010	35	0.101744
ylf	0-010	24	0.0956175
po	0-010	21	0.075
sg	0-010	44	0.0708535
ty	0-010	18	0.0642857
dka	0-010	12	0.0625
ms	0-010	17	0.0524691
cq	0-010	11	0.0371622
ve	0-010	4	0.0294118
isuyo	0-0100	42	0.242775
we	0-0100	28	0.202899
zg	0-0100	7	0.047619
le	0-0100	6	0.0454545
px	0-0100	26	0.0434783
sg	0-0100	26	0.041868
ylf	0-0100	9	0.0358566
xc	0-0100	33	0.0341615
xj	0-0100	5	0.027933
co	0-0100	2	0.010929
pixg	0-01000	34	0.213836
kdf	0-01000	17	0.184783
djh	0-01000	6	0.0425532
us	0-01000	16	0.0395062
wtz	0-01000	11	0.030303
nvvb	0-01000	8	0.0299625
bv	0-01000	10	0.021322
dka	0-01000	4	0.0208333
qj	0-01000	5	0.0182482
qlokt	0-01000	1	0.00980392
kdf	0-01001	20	0.217391

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
pixg	0-01001	32	0.201258
djh	0-01001	7	0.0496454
nvvb	0-01001	10	0.0374532
wtz	0-01001	13	0.0358127
po	0-01001	10	0.0357143
us	0-01001	14	0.0345679
pgk	0-01001	10	0.0290698
lho	0-01001	18	0.0235911
qlokt	0-01001	2	0.0196078
we	0-0101	35	0.253623
isuyo	0-0101	39	0.225434
ol	0-0101	18	0.0853081
djh	0-0101	11	0.0780142
pixg	0-0101	8	0.0503145
lho	0-0101	38	0.0498034
zg	0-0101	7	0.047619
le	0-0101	6	0.0454545
sg	0-0101	26	0.041868
kdf	0-0101	3	0.0326087
wtz	0-011	124	0.341598
ba	0-011	86	0.267913
djh	0-011	20	0.141844
us	0-011	44	0.108642
ve	0-011	14	0.102941
ylf	0-011	25	0.0996016
isuyo	0-011	16	0.0924855
zg	0-011	13	0.0884354
ol	0-011	16	0.0758294
kdf	0-011	6	0.0652174
ylf	0-1	59	0.23506
ba	0-1	16	0.0498442
ol	0-1	7	0.0331754
sz	0-1	7	0.0315315
nvvb	0-1	8	0.0299625

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
isuyo	0-1	5	0.0289017
kdf	0-1	2	0.0217391
bv	0-1	10	0.021322
px	0-1	11	0.0183946
we	0-1	2	0.0144928
pgk	0-10	127	0.369186
wtz	0-10	32	0.0881543
djh	0-10	12	0.0851064
dka	0-10	16	0.0833333
po	0-10	23	0.0821429
sg	0-10	46	0.0740741
hp	0-10	7	0.0636364
lho	0-10	46	0.0602883
bv	0-10	28	0.0597015
kdf	0-10	5	0.0543478
xj	0-100	53	0.296089
nvvb	0-100	15	0.0561798
qj	0-100	13	0.0474453
xq	0-100	25	0.0399361
kx	0-100	32	0.0376471
mtn	0-100	58	0.0363409
xc	0-100	34	0.0351967
wtz	0-100	12	0.0330579
we	0-100	4	0.0289855
cq	0-100	7	0.0236486
qlokt	0-1000	28	0.27451
lc	0-1000	39	0.18932
dt	0-1000	3	0.176471
pixg	0-1000	22	0.138365
tz	0-1000	26	0.104418
tg	0-1000	38	0.0865604
bv	0-1000	38	0.0810235
px	0-1000	40	0.0668896
ve	0-1000	9	0.0661765

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
ql	0-1000	69	0.0622744
qlokt	0-1001	34	0.333333
tz	0-1001	18	0.0722892
dt	0-1001	1	0.0588235
pixg	0-1001	9	0.0566038
fo	0-1001	15	0.0511945
sz	0-1001	11	0.0495495
bv	0-1001	22	0.0469083
jgy	0-1001	30	0.0466563
qj	0-1001	11	0.040146
ba	0-1001	9	0.0280374
le	0-10010	22	0.166667
cq	0-10010	30	0.101351
tz	0-10010	25	0.100402
dka	0-10010	16	0.0833333
ouy	0-10010	25	0.0684932
ol	0-10010	14	0.0663507
ql	0-10010	73	0.0658845
dt	0-10010	1	0.0588235
hp	0-10010	5	0.0454545
sz	0-10010	10	0.045045
le	0-10011	30	0.227273
ve	0-10011	30	0.220588
mtn	0-10011	150	0.093985
cq	0-10011	23	0.0777027
xq	0-10011	48	0.0766773
ol	0-10011	14	0.0663507
ql	0-10011	73	0.0658845
isuyo	0-10011	11	0.0635838
dt	0-10011	1	0.0588235
sz	0-10011	13	0.0585586
tz	0-101	63	0.253012
xj	0-101	44	0.24581
qj	0-101	37	0.135036

continued on next page ...

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
nvvb	0-101	33	0.123596
cq	0-101	30	0.101351
kx	0-101	66	0.0776471
ms	0-101	22	0.0679012
fo	0-101	19	0.0648464
mtsb	0-101	21	0.0630631
dt	0-101	1	0.0588235
xc	0-11	306	0.31677
pgk	0-11	58	0.168605
bv	0-11	61	0.130064
ouy	0-11	47	0.128767
cvf	0-11	7	0.118644
ql	0-11	130	0.117329
co	0-11	20	0.10929
cq	0-11	31	0.10473
mtn	0-11	163	0.10213
nvvb	0-11	27	0.101124
ty	0-110	87	0.310714
zg	0-110	37	0.251701
pgk	0-110	37	0.107558
qj	0-110	18	0.0656934
fo	0-110	15	0.0511945
sg	0-110	30	0.0483092
kx	0-110	40	0.0470588
mtn	0-110	70	0.0438596
hp	0-110	4	0.0363636
xj	0-110	6	0.0335196
zg	0-111	47	0.319728
ty	0-111	61	0.217857
ol	0-111	17	0.0805687
qj	0-111	13	0.0474453
ms	0-111	13	0.0401235
kx	0-111	34	0.04
fo	0-111	11	0.0375427

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
hp	0-111	4	0.0363636
xj	0-111	4	0.0223464
ve	0-111	3	0.0220588
djh	1-0	38	0.269504
ylf	1-0	5	0.0199203
ol	1-0	3	0.014218
pixg	1-0	2	0.0125786
mtsb	1-0	4	0.012012
mtn	1-0	19	0.0119048
ql	1-0	13	0.0117329
isuyo	1-0	2	0.0115607
lho	1-0	8	0.0104849
qlokt	1-0	1	0.00980392
sg	1-00	182	0.293076
we	1-00	17	0.123188
po	1-00	34	0.121429
ty	1-00	34	0.121429
wtz	1-00	44	0.121212
jgy	1-00	73	0.11353
nvvb	1-00	30	0.11236
ylf	1-00	23	0.0916335
hp	1-00	10	0.0909091
ol	1-00	18	0.0853081
lc	1-000	57	0.276699
tg	1-000	75	0.170843
sz	1-000	31	0.13964
xc	1-000	71	0.073499
lho	1-000	52	0.068152
fo	1-000	19	0.0648464
dka	1-000	12	0.0625
ql	1-000	67	0.0604693
co	1-000	10	0.0546448
mtn	1-000	87	0.0545113
tg	1-001	122	0.277904

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
lc	1-001	42	0.203883
px	1-001	73	0.122074
dt	1-001	2	0.117647
sz	1-001	24	0.108108
fo	1-001	30	0.102389
cq	1-001	30	0.101351
xq	1-001	56	0.0894569
ql	1-001	98	0.0884477
mtsb	1-001	24	0.0720721
px	1-01	201	0.33612
fo	1-01	36	0.122867
us	1-01	42	0.103704
we	1-01	14	0.101449
xj	1-01	17	0.0949721
pixg	1-01	13	0.081761
ylf	1-01	20	0.0796813
po	1-01	19	0.0678571
zg	1-01	9	0.0612245
co	1-01	9	0.0491803
ms	1-010	91	0.280864
dka	1-010	40	0.208333
zg	1-010	19	0.129252
ve	1-010	16	0.117647
qj	1-010	29	0.105839
jgy	1-010	66	0.102644
xq	1-010	63	0.100639
xc	1-010	93	0.0962733
sg	1-010	58	0.0933977
hp	1-010	9	0.0818182
dka	1-011	46	0.239583
ms	1-011	51	0.157407
ouy	1-011	32	0.0876712
sg	1-011	48	0.0772947
isuyo	1-011	13	0.0751445

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
lc	1-011	15	0.0728155
tz	1-011	17	0.0682731
pgk	1-011	23	0.0668605
xc	1-011	64	0.0662526
le	1-011	6	0.0454545
xq	1-1	169	0.269968
lc	1-1	23	0.11165
ms	1-1	24	0.0740741
tg	1-1	32	0.0728929
dt	1-1	1	0.0588235
qlokt	1-1	6	0.0588235
mtn	1-1	90	0.056391
bv	1-1	24	0.0511727
ve	1-1	6	0.0441176
jgy	1-1	26	0.0404355
mtn	2-0	441	0.276316
ve	2-0	25	0.183824
nvvb	2-0	42	0.157303
ty	2-0	40	0.142857
ol	2-0	30	0.14218
dka	2-0	26	0.135417
kdf	2-0	12	0.130435
mtsb	2-0	41	0.123123
ba	2-0	32	0.0996885
djh	2-0	14	0.0992908
lho	2-1	1	0.00131062
ql	2-1	1	0.000902527
cq	2-10	81	0.273649
xq	2-10	44	0.0702875
po	2-10	19	0.0678571
kdf	2-10	5	0.0543478
pgk	2-10	18	0.0523256
xc	2-10	43	0.0445135
sg	2-10	26	0.041868

continued on next page ...

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
mtn	2-10	62	0.0388471
ve	2-10	5	0.0367647
ql	2-10	36	0.032491
fo	2-100	87	0.296928
us	2-100	49	0.120988
tz	2-100	24	0.0963855
wtz	2-100	32	0.0881543
px	2-100	51	0.0852843
kx	2-100	67	0.0788235
xc	2-100	62	0.0641822
xq	2-100	37	0.0591054
dt	2-100	1	0.0588235
we	2-100	6	0.0434783
us	2-101	124	0.306173
lho	2-101	106	0.138925
ba	2-101	39	0.121495
xc	2-101	117	0.121118
qj	2-101	30	0.109489
pixg	2-101	17	0.106918
wtz	2-101	37	0.101928
tz	2-101	22	0.0883534
le	2-101	10	0.0757576
xq	2-101	38	0.0607029
mtsb	2-11	103	0.309309
tz	2-11	26	0.104418
ql	2-11	114	0.102888
we	2-11	14	0.101449
us	2-11	38	0.0938272
le	2-11	12	0.0909091
ouy	2-11	33	0.090411
pixg	2-11	14	0.0880503
co	2-11	9	0.0491803
ylf	2-11	11	0.0438247
po	2-110	97	0.346429

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
hp	2-110	13	0.118182
dt	2-110	2	0.117647
jgy	2-110	75	0.116641
px	2-110	67	0.11204
sg	2-110	69	0.111111
xc	2-110	103	0.106625
xj	2-110	18	0.100559
ouy	2-110	35	0.0958904
xq	2-110	59	0.0942492
ol	2-111	54	0.255924
kx	2-111	93	0.109412
kdf	2-111	10	0.108696
lho	2-111	79	0.103539
djh	2-111	14	0.0992908
pgk	2-111	31	0.0901163
qlokt	2-111	9	0.0882353
ms	2-111	28	0.0864198
cq	2-111	22	0.0743243
dt	2-111	1	0.0588235
ql	4-0	345	0.311372
we	4-0	18	0.130435
us	4-0	52	0.128395
le	4-0	15	0.113636
ouy	4-0	40	0.109589
ylf	4-0	27	0.10757
sg	4-0	66	0.10628
fo	4-0	31	0.105802
tg	4-0	43	0.0979499
mtn	4-0	149	0.0933584
klklv	4-1	1	1
pgk	4-1	5	0.0145349
lc	4-1	1	0.00485437
jgy	4-1	3	0.00466563
ylf	4-1	1	0.00398406

continued on next page . . .

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
po	4-1	1	0.00357143
cq	4-1	1	0.00337838
px	4-1	2	0.00334448
ba	4-1	1	0.00311526
wtz	4-1	1	0.00275482
nvvb	4-10	82	0.307116
dt	4-10	1	0.0588235
pixg	4-10	8	0.0503145
po	4-10	13	0.0464286
isuyo	4-10	6	0.0346821
cvf	4-10	2	0.0338983
wtz	4-10	11	0.030303
xj	4-10	5	0.027933
ylf	4-10	7	0.0278884
qlokt	4-10	2	0.0196078
bv	4-100	153	0.326226
mtsb	4-100	49	0.147147
le	4-100	15	0.113636
tz	4-100	28	0.11245
fo	4-100	30	0.102389
ms	4-100	32	0.0987654
ouy	4-100	35	0.0958904
kx	4-100	74	0.0870588
ba	4-100	26	0.0809969
kdf	4-100	7	0.076087
qj	4-101	104	0.379562
tg	4-101	49	0.111617
px	4-101	66	0.110368
cvf	4-101	6	0.101695
lho	4-101	73	0.095675
mtn	4-101	151	0.0946115
isuyo	4-101	16	0.0924855
mtsb	4-101	29	0.0870871
co	4-101	12	0.0655738

continued on next page ...

Example of an agent's complete lexicon

Signal	Meaning	Usage	Conf. Prob
kdf	4-101	5	0.0543478
lho	4-11	232	0.304063
jgy	4-11	83	0.129082
qlokt	4-11	13	0.127451
hp	4-11	13	0.118182
bv	4-11	51	0.108742
kx	4-11	86	0.101176
sz	4-11	22	0.0990991
ylf	4-11	24	0.0956175
xj	4-11	17	0.0949721
cvf	4-11	5	0.0847458

Table E.2: Example of an agent's active lexicon

Signal	Meaning	Usage	Conf. Prob
klklv	4-1	1	1
qj	4-101	104	0.379562
pgk	0-10	127	0.369186
jgy	0-01	225	0.349922
ba	0-010	112	0.34891
po	2-110	97	0.346429
wtz	0-011	124	0.341598
px	1-01	201	0.33612
qlokt	0-1001	34	0.333333
bv	4-100	153	0.326226
ouy	0-0	118	0.323288
zg	0-111	47	0.319728
xc	0-11	306	0.31677
co	0-000	57	0.311475
ql	4-0	345	0.311372
ty	0-110	87	0.310714
mtsb	2-11	103	0.309309
nvvb	4-10	82	0.307116
us	2-101	124	0.306173
lho	4-11	232	0.304063
kx	0-00	254	0.298824
fo	2-100	87	0.296928
xj	0-100	53	0.296089
sg	1-00	182	0.293076
ms	1-010	91	0.280864
tg	1-001	122	0.277904
lc	1-000	57	0.276699
mtn	2-0	441	0.276316
cq	2-10	81	0.273649
cvf	0-0001	16	0.271186
xq	1-1	169	0.269968
djh	1-0	38	0.269504
sz	0-0000	59	0.265766

continued on next page . . .

Example of an agent's active lexicon

Signal	Meaning	Usage	Conf. Prob
ol	2-111	54	0.255924
we	0-0101	35	0.253623
tz	0-101	63	0.253012
isuyo	0-0100	42	0.242775
dka	1-011	46	0.239583
ylf	0-1	59	0.23506
le	0-10011	30	0.227273
kdf	0-01001	20	0.217391
pixg	0-01000	34	0.213836
hp	0-001	22	0.2
dt	0-1000	3	0.176471

APPENDIX F

Published Papers

This appendix contains articles which were published, accepted for publication, or under review prior to the completion of this thesis, and consists of a journal article (Smith (2003a)), two articles published in collections on Artificial Intelligence (Smith (2001, 2003b)), and one article which will appear in a collection on the Evolution of Language (Smith (forthcoming)). Details of all these articles are given below:

Smith, A. D. M. (2001). Establishing communication systems without explicit meaning transmission. in J. Kelemen and P. Sosík (Eds.), *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life* (pp. 381–390). Heidelberg: Springer-Verlag.

Smith, A. D. M. (2003a). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life* 9(2) 175–190.

Smith, A. D. M. (2003b). Semantic generalisation and the inference of meaning. in W. Banzhaf, T. Christaller, J. Ziegler, P. Dittrich and J. T. Kim (Eds.), *Advances in Artificial Life: Proceedings of the 7th European Conference on Artificial Life* (pp. 499-506). Heidelberg: Springer-Verlag.

Smith, A. D. M. (forthcoming). Mutual exclusivity: Communicative success despite conceptual divergence. in M. Tallerman (Ed.), *Evolutionary prerequisites for language*. Oxford: Oxford University Press.

Smith, A. D. M. (2001). Establishing Communication Systems without Explicit Meaning Transmission. in J. Kelemen and P. Sosík (Eds). *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life* (pp. 381–390). Heidelberg:Springer-Verlag.

Establishing Communication Systems without Explicit Meaning Transmission

Andrew D.M. Smith

Language Evolution and Computation Research Unit,
Department of Theoretical and Applied Lingustics, University of Edinburgh, UK
andrew@ling.ed.ac.uk

Abstract. This paper investigates the development of experience-based meaning creation and explores the problem of establishing successful communication-systems in a population of agents. The aim of the work is to investigate how such systems can develop, without reliance on phenomena not found in actual human language learning, such as the explicit transmission of meaning or the provision of reliable error feedback to guide learning. Agents develop individual, distinct meaning structures, and although they can communicate despite this, communicative success is closely related to the proportion of shared lexicalised meaning, and the communicative systems have a large degree of redundant synonymy.

1 Introduction

There is a growing body of literature in which investigations into the evolution of language[1] are carried out by computer simulation [8,1,4]. For most of these researchers, the evolution of language is regarded as essentially being equivalent to the evolution of syntax, because the use of syntactic structure is seen as the main difference between animal and human communication systems. For example, vervet monkeys have a well-known communication system which allows them to distinguish different predators [5], but they do not combine their signals to convey complex meanings. Kirby [9] has shown that the simple ability to create general rules, by taking advantage of coincidental correspondences between parts of utterances and parts of meanings, can result in the emergence of syntax, as general rules generate more utterances than idiosyncratic rules, and are therefore replicated in greater numbers in following generations. Similar accounts [2,10] also show syntax emerging as a consequence of the recognition and coding of regularities between signals and meanings.

Nehaniv [14] has pointed out, however, that syntax only develops successfully from unstructured signals because the signals are coupled with meanings which are already structured, and it is no coincidence that the emergent syntactic structure parallels the pre-existing semantic structure. In these simulations,

[1] This field is concerned not with the evolution of particular languages, such as English, from their ancestor languages, but rather with the general capacity, apparently unique to humans, for using infinitely expressive communication systems [13].

J. Kelemen and P. Sosík (Eds.): ECAL 2001, LNAI 2159, pp. 381–390, 2001.

the meanings are also explicitly part of the linguistic transfer from speaker to hearer, therefore obviating the critical problem, exemplified by Quine [17](p. 29–30), of how a learner determines the meaning which a signal intends to convey. Furthermore, attempts to develop learnt communication systems frequently involve some sort of reinforcement learning process [20,6], which has the primary role in guiding the learning mechanism. Oliphant [15] points out, however, that such error signals, which work well on an evolutionary timescale, are less useful over an individual's lifetime where failure might mean immediate death, and indeed even the very existence of reliable error signals is questioned by many authors on child language acquisition [3].

If we try to define the meaning of a word, we find ourselves caught in a kind of lexical web, where words can only be defined by their relationship to other words, and in terms of other words. There is no obvious way of entering this web, unless at least some words are grounded in reality [7], such that they can be used to point out actions and objects in the real world. It is reasonably uncontroversial to say that meanings must capture patterns of categorisation (whether categories are defined in classical terms of shared features or prototypes [21]) which enable us to state, for instance, which things are *rabbits* and which are not. Furthermore, meanings are not innate, but are created anew in each language learner, who creates an individual system of meaning based on their experiences [3].

Our aim is to model, in a population of agents, the creation of meanings by explicit categorisation, and then to investigate the spread of meanings through the population, without the meanings themselves being transferred between agents, and without any error signals to reinforce the learning process.

2 Meaning Creation by Object Discrimination

In order to develop a model of independent, grounded meaning creation, we establish a simple world of agents and objects, similar to that described by Steels [19], in which the objects can be described in terms of their features[2], which are intrinsically meaningless, but which can be thought of in terms of more imaginable language-like features such as *colour, height* or *smell*. The agents in the model world interact with the objects by using *sensory channels*, which are sensitive to the corresponding features of objects, and can detect whether a particular value of a feature falls between two bounds. Initially, the channels can only detect that a value falls between 0.0 and 1.0, but the agents have the power to split the sensitivity range of a channel into two discrete segments, resulting in a *discrimination tree* [20]. The nodes of a discrimination tree can be considered categories or *meanings*,[3] as seen in the sensory channel in figure 1, which has been refined twice, and has four new meanings.

[2] Feature values are represented as pseudo-randomly generated real numbers which are normalised to lie between 0.0 and 1.0

[3] Meanings are given in the notation *sc-path*, where *sc* identifies the sensory channel, and *path* traces the path from the tree root to the node in question, where 0 signifies a lower branch and 1 an upper branch.

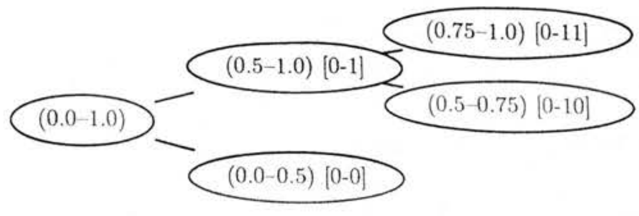


Fig. 1. A discrimination tree (channel 0) which has been refined twice. Each node shows the bounds between which it is sensitive, and the meaning to which it corresponds (following Steels).

In order to provide a framework for the unguided refinement of the sensory channels based on observation, we follow Steels [19] in using discrimination games, in which an agent attempts to distinguish one object from a larger set of objects. Each game proceeds as follows:

1. An agent considers a random set of objects (the *context*), one of which is chosen at random to be distinguished from the others and is called the *topic*.
2. The agent investigates all its sensory channels to categorise the objects.
3. If the topic is uniquely identified by any category, the game succeeds.
4. If the game fails, the agent refines a randomly-chosen sensory channel.

Object	Categories/Meanings		
A	0-0	1-00	2-111
B	0-11	1-1	2-110
C	0-0	1-1	2-111
D	0-10	1-01	2-10

The above table shows an agent categorising objects as part of a discrimination game. The agent has four objects A-D, and has categorised them with three sensory channels. If the aim of this game is to discriminate B from the context ACD, then the game can succeed, as both $0-11$ and $2-110$ are possible distinguishing categories. On the other hand, if the aim is to distinguish C from the context ABD, then the game will fail, as none of the categories which C falls into distinguish it from all the other objects. Failure triggers the refinement of a random channel, creating more detailed categories, which *may* be useful in future games. Over time, the agents develop their sensory channels such that the discrimination games nearly always succeed, though the extent to which an individual channel is refined depends on the number of channels which the agent has: the more channels, the fewer refinements on each are necessary.

Figure 2 shows the idiosyncratic meaning representations of two agents in the same world. The first agent has developed the first three channels to a greater extent than the second agent, who in turn has developed the fourth and fifth channels more extensively. It is helpful to quantify the amount of difference between two trees t_1 and t_2, which we can do by averaging the proportion of

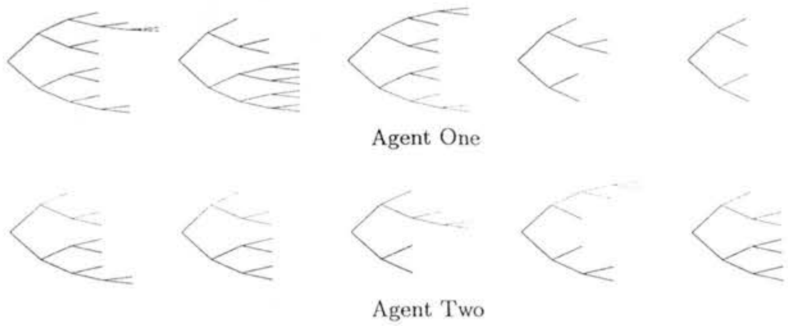


Fig. 2. Two agents each have five sensory channels, with which they construct different representations of the same world.

nodes in tree t_1 which are also in t_2, and the proportion in t_2 which are also in t_1. Averaging over all the trees in figure 2, the two meaning representations have a *meaning similarity* measure of 75%. It is important to note that both agents are successful in the discrimination games, and so their representations are equally good descriptions of their world. This model, then, satisfies one of our goals, namely that the agents are not given innate meanings, but can create inventories of basic concepts individually, based on their own experiences.

3 Communication

The next step is to investigate whether the agents can communicate with each other, using the meanings they have constructed. Clearly the agents must be able to use some sort of signals, and so they are endowed with the ability to create signals from random strings of letters, and to express and understand these signals without error. In addition, they maintain a dynamic lexicon of associations between signals and meanings, which develops as they participate in the experiments, and which they use in order to make decisions about their communicative behaviour. *Communicative success* occurs if the speaker and hearer are both referring to the same object, but it is not necessary for them to use the same meaning to do so.

A *communicative episode* is played between two agents chosen at random, the *speaker* and the *hearer*. Figure 3 shows a model of the speaker's role, which begins with a discrimination game, in which meanings which can distinguish the topic (filled circle) from the rest of the context (dashed area) are collated. One of these meanings is chosen at random and then looked up in the speaker's lexicon. If the speaker cannot find a word for the meaning it is trying to convey, then it creates a random string of letters and stores this in its lexicon with the required meaning. Having obtained a word to convey the meaning, the speaker utters the word, and the focus passes to the hearer, who receives the word, and can observe the context in which it was uttered, shown in figure 4.

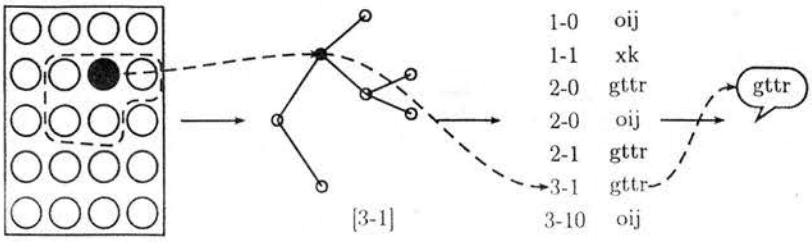


Fig. 3. A communicative episode begins with an agent chosen at random to be the speaker, who finds a meaning to distinguish the topic from the context, and utters a word to convey this meaning.

The word is decoded via the hearer's lexicon into a meaning, and the hearer then establishes which object in the context (if any) is uniquely identified by the meaning it has chosen. If the referent (object) identified by the hearer corresponds to the speaker's original topic, then the communication episode succeeds. The success or failure of a communication game has no effect on the internalised representations of either agent. This model of communication conforms to our initial assumptions, as the internal meanings are explicitly *not* transmitted with the signals, and the agents do not receive feedback from each other about the success of their communicative or learning processes.

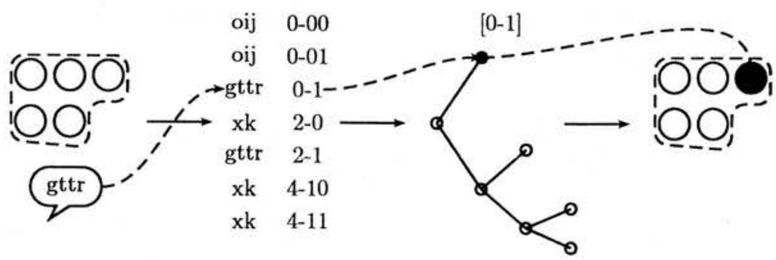


Fig. 4. The communicative episode continues with the hearer, who, given the context, decodes the word into a meaning which identifies an object.

4 The Lexicon

The mappings from meaning to signal and vice-versa are at the heart of the communication process, and are handled via a lexicon, which stores associations between meanings and signals, as well as a count of how often the signal-meaning pair has been used (either uttered as a speaker or understood as a hearer), and a confidence probability, which represents the agent's confidence in the association between the signal and the meaning.

The confidence probability of the signal-meaning pair, consisting of signal s and meaning m, represents the history of all associations between words and meanings an agent has ever made, and is defined as the proportion of the times s has been used in which it has been associated with m, or $\frac{Usage(s,m)}{\sum_1^l Usage(s,l)}$ where l is the number of entries in the lexicon. A short extract from an example lexicon is given below, only showing the entries for two of the signals (*gttr* and *oij*), and the meanings associated with them.

Signal	Meaning	Usage	Conf. Prob.
gttr	0-0	1	0.083
gttr	0-1	2	0.167
gttr	0-11	1	0.083
oij	1-0	9	0.600
gttr	2-0	4	0.333
oij	2-0	6	0.400
gttr	2-1	1	0.083
gttr	3-1	2	0.167
gttr	4-00	1	0.083

How does the speaker decide which signal to choose, when it is trying to express a particular meaning (say $2-0$)? Given the lexicon above, the signal *oij* would seem a reasonable choice for two reasons: it has been associated with $2-0$ on six occasions, compared to *gttr*'s four, and the agent is more confident in the association with *oij* (0.4) than *gttr* (0.33). However, Oliphant and Batali [16] have demonstrated an ideal strategy for achieving an accurate communication system, known as *obverter*, where the speaker chooses words which he knows the hearer will understand. Unfortunately, true obverter learning assumes that the speaker can read the lexicons of the other members of the population, to calculate the optimal signal to use for any meaning. Such mind-reading is not only unrealistic, but even avoids the need for communication at all, and so an alternative is needed. It seems reasonable to assume that the only lexicon the speaker has access to is its own, and so we assume that the speaker uses this as an approximation to that of the hearer. Instead of explicitly choosing the word that the hearer will understand, the speaker chooses the word that *it* would be most likely to understand if it was the hearer. Returning to the lexicon above, we can see that although *oij* has been associated with the meaning $2-0$ on more occasions than *gttr*, if heard, it would actually be interpreted as $1-0$ (because $1-0$ is the meaning which maximises the confidence probability for *oij*), whereas *gttr* would be interpreted with the correct $2-0$ meaning.

Interestingly, the agent would not find a word from its lexicon to express many meanings which do have some associations (e.g. $0-0$, $3-1$ etc.). One of the outcomes of obverter learning is the avoidance of ambiguity, so we find that, at any one time, each word in the lexicon is only used with one meaning, although the particular meaning can of course change as the associations in the lexicon are updated. This means that, although there are eight meanings in the lexicon extract, only two of them are actually used by the speaker, and so only these can be regarded as being truly *lexicalised.*

We have seen how the speaker tries to second-guess the hearer and chooses words which are likely to be understood before uttering them, but a greater problem is faced by the hearer in understanding the meaning which is being conveyed. On hearing a signal, the hearer's only guide in determining the intended meaning is the observation of the context (which of course includes the target topic object). From this, the hearer constructs a list of all the possible meanings, that is, *all* meanings which categorise only one of the objects in the context. All these possible meanings are equally plausible, so the hearer associates each of them with the signal in its lexicon, adjusting its confidence probability for each accordingly. Over time, the interpretation of each word will tend to the speaker's intended meaning, if the two agents have identical meaning structures [18].

5 Results

The meaning structures constructed by the agents in our model world, however, are of course not only not identical, but also change over time. Under these circumstances, is it possible for the agents to communicate? Figure 5 (left) shows

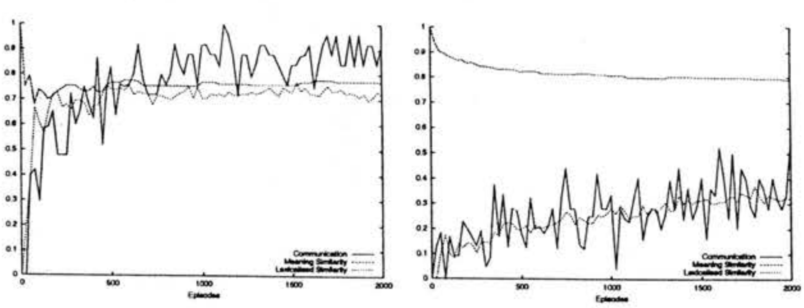


Fig. 5. Communicative success, meaning similarity, and lexicalised similarity for a population of two agents and 100 objects. Each discrimination game is played with a context size of five objects. The number of sensory channels available to each agent is five (left) and 100 (right).

that communication is successful a large percentage of the time, although it is not optimal, and does not appear to increase significantly after the initial rise to around 90%. The similarity of the agents' meaning structure drops initially, as the agents refine their sensory channels individually and separately, and then does not change significantly. This occurs because the pressure to develop meaning structure comes only from failure in discrimination games, and after an initial flurry, the agents all have sufficiently detailed meanings to succeed in nearly all discrimination games. Once this state is achieved, the communication rate stops improving and remains fairly constant. If the number of sensory channels available is increased substantially (figure 5: right), a similar result is found, except that the rate at which communication stops improving is much lower. It can also be seen that the communication success rate is closely parallelled in both cases by the *lexicalised similarity* of the agents, which is defined in the same way as meaning similarity (see section 3), but only taking into account tree nodes which are lexicalised.

An interesting phenomenon which occurs in these kind of simulations is the large amount of synonymy which pertains in the lexicons, where more than one word is interpreted with the same meaning. As an example, after 1000 communicative episodes, two agents have the meaning structures shown in figure 6. Attached to each node on the discrimination trees is the number of words which

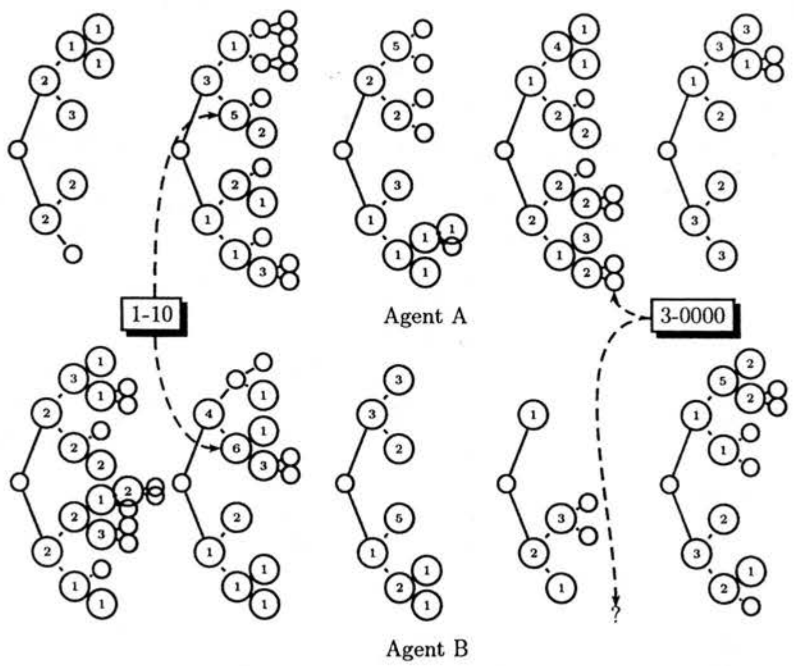


Fig. 6. Two agents each have five discrimination trees numbered 0-4. Each lexicalised node is marked with the number of words which would be interpreted as that meaning.

this agent would interpret as the meaning denoted by the node, or the number of *synonyms* attached to the meaning. For instance, we can see that there are five words which would each be interpreted by agent A as $1-10$, and six which would be interpreted as this by agent B. Further inspection (not shown) indicates that four of these synonyms have been lexicalised by both agents, suggesting a high level of redundancy, which is caused by *meaning drift*.

The interpretation of a word, of course, changes over time as the agents develop their experience of the word's use. Words are only created when an agent wants to express a meaning which isn't lexicalised. For example, in figure 6, agent A might wish to express the meaning $3 - 0000$, but it does not have a word which it would interpret correctly, so it creates a new word *ujszo*. Agent B hears the new word, and creates a list of possible meanings. This list, however, cannot include A's meaning $3 - 0000$, because B's meaning structure does not contain this meaning, and so B will lexicalise *ujszo* with a different meaning. Over time, B's preferred meaning is likely to be a more general meaning, which is shared by A.[4] There is now a difference of opinion over the meaning of *ujszo*, but crucially, agent A can continue to associate it with B's meaning, while B cannot associate it with A's original meaning. A's association between *ujszo* and the shared meaning gradually increases, until it eventually exceeds that of the original meaning. Both agents will now use *ujszo* for the more general meaning: the word's meaning has drifted. As a direct consequence, A no longer has a word with which it can express the meaning $3 - 0000$. If it does need to convey this meaning, it must create another new word, and the cycle begins again.

Meaning drift is an inevitable characteristic of systems in which the agents' conceptual systems are not the same, if there are an unlimited number of signals, and there is little pressure to modify meaning structure. Inducing the meanings of words from context inevitably biases the meanings towards those meanings which are more general, and shared by the agents. Words which refer to specific meanings which are not shared will see their meanings drift to those which are shared, resulting in a large number of synonyms for the shared meanings, and few, if any, words at all for the agent-specific meanings.

6 Discussion

We have developed a world in which agents can communicate about their environment, without explicitly transferring meanings, without knowing exactly what the speaker is referring to, and without providing the learner with any feedback about communicative success, all criteria motivated by research into how human children acquire language [3]. Although communication can succeed in cases where agents refer to the same object with different meanings, the overall success of communication seems to be directly related to the amount of shared meaning structure in the agents. The communication system has a great deal of synonymy, caused by the differences in meaning structure and the unlimited number of possible signals. Work is under way to extend the model, focusing on ways to reduce synonymy, for instance by implementing the *principle of contrast* [12], and to investigate the effects of specific biases in meaning induction,

[4] Because general meanings are created before more specific meanings on the discrimination trees, they are more likely to occur in both agents' meaning structures.

such as the *shape bias* [11]. It is claimed that such biases explain the learning of meanings [3], and this work will go some way to showing where these claims are feasible.

References

[1] J. Batali. Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks and P. Maes, editors, *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pages 160–171. Addison-Wesley, Redwood City, CA, 1994.

[2] J. Batali. The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, Cambridge, 2001.

[3] Paul Bloom. *How Children Learn the Meanings of Words*. MIT Press, 2000.

[4] E. Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296, 2000.

[5] D. Cheney and R. Seyfarth. *How Monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press, Chicago, IL, 1990.

[6] Edwin D. de Jong. *Autonomous Formation of Concepts and Communication*. PhD thesis, Vrije Universiteit Brussel, 2000.

[7] S. Harnad. The symbol grounding problem. *Physica*, D 42:335–346, 1990.

[8] James R. Hurford. The evolution of critical period for language acquisition. *Cognition*, 40:159–201, 1991.

[9] S. Kirby. Learning, bottlenecks and the evolution of recursive syntax. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, Cambridge, 2001.

[10] V. Kvasnička and J. Pospíchal. An emergence of coordinated communication in populations of agents. *Artificial Life*, 5(4):319–342, 1999.

[11] Barbara Landau, Linda B. Smith, and Susan S. Jones. The importance of shape in early lexical learning. *Cognitive Development*, 3:299–321, 1988.

[12] E. M. Markman. *Categorization and naming in children*. MIT Press, 1989.

[13] John Maynard Smith and Eörs Szathmáry. *The major transitions in evolution*. Oxford University Press, 1995.

[14] Chrystopher L. Nehaniv. The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In *Proceedings of the AISB Symposium: Starting from Society — the application of social analogies to computational systems, 19–20 April 200*, pages 73–84, 2000.

[15] M. Oliphant. The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, 7(3/4):371–384, 1999.

[16] M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, 11(1), 1997.

[17] W. V. Quine. *Word and Object*. MIT Press, 1960.

[18] A. D. M. Smith. The role of shared meaning structure in communication. Paper given at the 7^{th} Human Behaviour & Evolution Society Conference, London, 2001.

[19] L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the international conference on multi-agent systems*. MIT Press, 1996.

[20] L. Steels. *The Talking Heads Experiment*, volume I. Words and Meanings. Laboratorium, Antwerpen, 1999. Special pre-edition.

[21] J. R. Taylor. *Linguistic Categorization*. Oxford University Press, 1995.

Smith, A. D. M. (2003a). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life* 9(2) 175–190.

Intelligent Meaning Creation in a Clumpy World Helps Communication

Andrew D. M. Smith
Language Evolution and Computation Research Unit
Theoretical and Applied Linguistics
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Adam Ferguson Building
40 George Square
Edinburgh, EH8 9LL
United Kingdom
andrew@ling.ed.ac.uk

Abstract This article investigates the problem of how language learners decipher what words mean. In many recent models of language evolution, agents are provided with innate meanings a priori and explicitly transfer them to each other as part of the communication process. By contrast, I investigate how successful communication systems can emerge without innate or transferable meanings, and show that this is dependent on the agents developing highly synchronized conceptual systems. I present experiments with various cognitive, communicative, and environmental factors which affect the likelihood of agents achieving meaning synchronization and demonstrate that an intelligent meaning creation strategy in a clumpy world leads to the highest level of meaning similarity between agents.

Keywords
Meaning similarity, meaning creation, communication, language evolution

1 Introduction

Attempts to explain the particular structure of language often appeal to a "conventional neo-Darwinian process" [21], whereby humans have evolved an innate, genetically encoded language device in the brain which is specifically tailored to the acquisition and maintenance of language [5]. More recently, however, researchers have begun to develop models which emphasize the repeated process of language learning and use it as the driving force behind the emergence of linguistic structures. For example, Kirby [12] explores in detail how certain language universals [9] can be explained elegantly by focusing on how processing complexity affects the transmission of language.

Much recent work in the field of language evolution has focused on the evolution of syntactic structure as the crucial event which marks both the genesis of language and the defining criterion which separates it from animal communication systems. Kirby [13], for example, demonstrates that syntax can arise from unstructured communication systems by creating generalized rules from the analysis of signal-meaning pairs, and Brighton [4] shows that pressures such as the poverty of the stimulus [5] lead to the emergence of syntactic structure when the process of language production and learning is repeated over generations.

There are, however, some major problems with the assumptions behind simulations such as these. Firstly, syntax develops only because signals in the simulations are coupled to pre-existing, innate, structured meanings, and so it is no surprise to find that the structure of the emergent syntax directly parallels that of the predefined semantics, as discussed by Nehaniv [19]. Explanations of the origin of these meanings, and of how they become associated with signals, are conspicuously absent. Secondly, communication consists of the simultaneous transfer of signals and meanings; thus the simulations ignore one of the most crucial features of real language acquisition, namely that meanings are *not* transferred with words, and yet learners do manage to infer meanings and associate words with them. Thirdly, the simulations rely on variants of reinforcement learning to guide the agents [26], although the existence of reliable error signals in language learning is widely rejected [3]. In contrast, I argue that constructing meanings and learning which of them are most relevant is a crucial part of the language learning process which should not be overlooked.

The article is divided into six main parts. In Section 2, I discuss the assumption of explicit meaning transfer and its implications for models of communication and learning. In Section 3, I report details of the model of meaning creation and communication, describing how the problem of explicit meaning transfer can be overcome. In Section 4, I show the importance of meaning similarity for the emergence of successful communicative systems, and describe a baseline for meaning similarity. Finally, in Sections 5–7, I investigate how cognitive biases, communicative biases, and environmental factors such as the agents' experience and the structure of the world affect levels of meaning similarity, and therefore levels of successful communication.

2 Explicit Meaning Transfer

Kirby [13] and Batali [2] have shown separately how the simple ability to create general rules, by taking advantage of coincidental correspondences between parts of utterances and parts of meanings, can result in the emergence of a compositional, syntactic communication system. In a nutshell, this occurs when the agents are subject to pressures which limit their exposure to the language, such as the poverty of the stimulus; general rules can generate more utterances than idiosyncratic rules, are more likely to be encountered, and are therefore replicated in greater numbers in following generations. I have already noted, however, that the successful emergence of syntax in these models is dependent on the signals being coupled to structured meanings. The structure of the meanings is assumed by the model, and it is not coincidental that the syntactic structure which emerges parallels exactly the pre-existing semantic structure.

At the heart of any kind of communication system is what constitutes observable behavior during linguistic transfer, or what is actually transmitted between speakers and hearers. In Figure 1, which represents the linguistic transfer in a standard model, we can see that the speaker (on the left of the picture) utters a signal "zknvrt," but that simultaneously, the meaning in the speaker's brain (represented by three apples) is transferred directly to the hearer's brain. The hearer learns the association between signal and meaning, and crucially, it knows that ***this association*** is appropriate to make because the signal and meaning are explicitly linked in each communicative episode.

This kind of model of associative learning sidesteps one of the most important and difficult problems facing researchers into the acquisition of language, namely Quine's [22] famous *gavagai* problem of determining the meaning of an unfamiliar word from a set which is, in principle, infinite. The consequences of this idealization of the learning process are considerable, not least because if meanings are explicitly and accurately transferable by telepathy as in Figure 1, then the signals are not being used to convey meaning. If the signals do not convey meaning, then their role in the model is far from obvious. In fact, we can see that the inclusion of signals in the model is a complicating factor, and yet removing them brings us uncomfortably close to creating a model which bears very little resemblance to a languagelike communication system. We are left, therefore, with the conclusion that meanings *cannot* be explicitly transferred, but must instead be inferred by the hearer from the signal and the context in which they are heard.

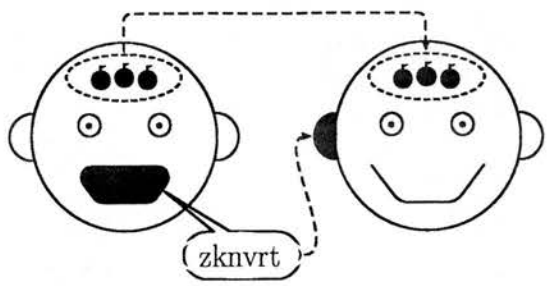


Figure 1. A communicative episode which consists of the explicit transfer of both a signal "zknvrt" and a meaning "three apples" from speaker to hearer.

So how, then, does a hearer know which meaning to associate with a signal, and where do the private meanings it uses come from? Firstly, if it is assumed that meanings are not transferable, then the agents must be able at least to infer them from elsewhere. I assume that the obvious, and most general, source for this is the world around the agent, or the environment in which it is placed. This in turn suggests that at least some of the meanings which agents talk about are used to refer to objects and events which actually happen in the environment. Binding the subjects of communication to events in the agents' world means that the agents' meanings are grounded in the world [10].

It is worth noting that the need to infer meanings from the environment has interesting implications for models such as those described by Kirby [13] and Batali [2]. These models contain no environment, and indeed nothing accessible and external to the agents, so the "meanings" used must necessarily be abstract, predefined tokens. Because they can have no reference (cannot identify any thing in the world), they cannot be inferred, and so can only be communicated through explicit transfer. In order to avoid explicit meaning transfer, therefore, there must be some kind of external world for the agents to experience in the model.

The existence of an external world in itself, however, does not mean that the problem of explicit meaning transfer is automatically avoided; for this there must be at least three separate levels of representation in the model: the external, public world, a private, agent-specific internal semantic representation, and a set of signals, which can again be publicly observed. The mappings between the public and private sections of the model must be specific to each agent and unobservable to the others; otherwise the private representations become public, making the signals unnecessary.

In Hutchins and Hazlehurst's famous neural network model of the development of a shared vocabulary [11], for instance, there is an external world made up of events, or "scenes." These scenes, however, are themselves used as the meanings for which the agents learn signals; although they are not explicitly transferred, they are publicly accessible in the communication process, and there is therefore no level of the model which is private to each agent. Brighton [4], too, presents a model with an external world made up of communicatively relevant situations. But although the environment is defined as the source of the meanings used by the agents, this relationship plays no role in the simulations; the agents never interact with the environment, and the mapping from environment to meanings is predetermined and identical for all agents. Again, there is no private level in the model, and the environment is effectively merely a complicating factor in the simulation.

Secondly, there are two possible explanations for how the agents come to have meanings which refer to things: either the meanings are innate, and have somehow evolved biologically, or they are created by the agents themselves, as a result of their interactions with the environment. Innate meanings are not inherently implausible, and they are used as a simplification in many models of aspects of language evolution (see for instance Arita and Koyama [1]), but they seem in reality to require either that the number of meanings useful to the agents be small and fixed, or that the world in which the agents exist be very stable and unchanging. If the world is dynamic, then the agents may have evolved innate meanings for something that was useful to their ancestors, but these may not be of use to them now. In practice, then, I assume that it is more reasonable to assume that the agents create meanings *de novo* in each generation, based on empirical testing of their environment, to discover which distinctions are communicatively relevant.

This paper, therefore, departs from previous accounts, which assume that language learning is equivalent to learning a mapping between signals and predefined meanings. Instead, I argue that there are at least three necessary levels of representation: a public environment, a private semantic representation, and public signals. Language learning involves the empirical creation of private meanings based on the environment, learning which of these meanings are relevant, and learning the mapping between signals and the relevant meanings which underpins communication.

3 Details of the Model

3.1 Meaning Creation

My model of independent, grounded meaning creation is based on that described by Steels [25]. I establish a simple world made up of a number of objects, which can be described in terms of the values of their *features*. In the results reported here, the world contains twenty objects unless otherwise specified. Feature values in the model are real numbers, pseudo-randomly generated in the range [0,1]. These features are abstract and do not have any specified meaning in the model, but can be profitably thought of in terms of perceptual features such as smell or color. The agents in the world interact with the objects using *sensory channels*. They have the same number of sensory channels as the objects have features, and there is a one-to-one mapping between channels and features. Sensory channels are sensitive to the feature values, and in particular can detect whether a particular feature value falls between two bounds. Meaning creation happens by splitting the sensitivity range of a channel into two discrete segments, resulting in two separate categories, or *meanings*, each sensitive to half the original range. After repeated splitting or refinement, we can represent the semantic structure on a dendrogram, as shown in Figure 2, where the nodes on the tree represent the meanings.

The agents interact with their environment through *discrimination games* [25], in which they try to distinguish one particular randomly chosen object from a context of five randomly chosen objects through the following algorithm:

- The agent investigates all its sensory channels to categorize all the objects in the context.
- If the target object is uniquely identified by any single category, then this meaning is called the *discriminatory meaning* and the game succeeds.
- If the game fails, the agent refines a randomly chosen sensory channel.

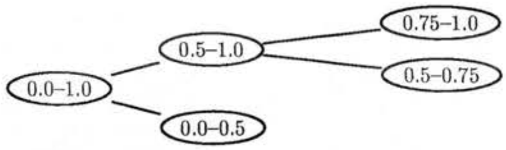


Figure 2. Meanings represented on a Steelsian dendrogram, which has been refined twice. Each node on the tree shows the bounds between which it is sensitive.

Table 1. The categorization of objects during a discrimination game. Meanings are given in the notation c–p, where c identifies the sensory channel and p traces the path along the discrimination tree from the root to the node in question, with 0 signifying a lower branch and 1 an upper branch.

Object	Categories/Meanings		
	Channel 0	1	2
A	0–0	1–00	2–111
B	0–11	1–1	2–110
C	0–0	1–1	2–111
D	0–10	1–01	2–10
E	0–10	1–00	2–0

Table 1 shows an agent's categorization of objects during a discrimination game; the agent is investigating five objects, and has three sensory channels on which the objects are being categorized. If the aim of this game is to discriminate *B* from the context *ACDE*, then the game can succeed, as both 0–11 and 2–110 are possible discriminatory meanings. On the other hand, if the aim is to distinguish *C* from the context *ABDE*, then the game will fail, as there is no single category into which *C* falls which distinguishes it from all the other objects.

Failure in such a discrimination game triggers the refinement of a randomly chosen sensory channel, and therefore the creation of another level of conceptual structure in the agent. Because the sensory channel is chosen randomly, the newly created meanings may be, but are not necessarily, useful for future discrimination games. Given enough discrimination games in a static world, the agents will always develop a successful conceptual structure, although the precise details of this structure are of course not fixed, and will vary between agents and between runs of the simulation.

This semantic representation has an obvious hierarchical structure, allowing the immediate use of real semantic sense relationships such as hyponymy and antonymy to be investigated, which are not readily available in other representations. Meanings nearer the root of the tree are clearly more general than those nearer the leaves of the tree, which are more specific. Concept creation is clearly directly driven by the agents' interactions with their world, so that the meanings are not imposed from outside. The agents, therefore, have a mechanism for constructing concepts which is grounded in the environment, is based on experience, creates meanings which are useful to the agents in allowing them to discriminate between the objects they find, and results in conceptual structure which can be measured and compared. We quantify the similarity of two agents' meaning structures by averaging the similarity of the particular discrimination trees built on each of their sensory channels in turn. In greater detail, if $k(t, u)$ is the number of nodes which trees t and u have in common, and $n(t)$ is the total number of nodes on tree t, then we describe the similarity between any two trees t

and u using the following formula:

$$\tau(t, u) = \frac{1}{2}\left(\frac{k(t, u)}{n(t)} + \frac{k(t, u)}{n(u)}\right) \tag{1}$$

We can use this general measure of tree similarity τ to develop an overall measure of *meaning similarity* σ between two agents, by averaging over all their sensory channels. If $a_i t_j$ identifies channel j on agent i, and each agent has c sensory channels, then the meaning similarity σ between agents a_1 and a_2 is defined as follows:

$$\sigma(a_1, a_2) = \frac{1}{c}\sum_{i=0}^{c-1} \tau(a_1 t_i, a_2 t_i) \tag{2}$$

If two agents a_1 and a_2 have identical conceptual structures, where $\sigma(a_1, a_2) = 1$, then we refer to their meanings as being *synchronized*.

3.2 Communication

In this section, I extend the meaning creation model to investigate whether the agents can communicate with each other, using the meanings they have constructed. In order to simulate communication between the agents, I endow them with the ability to create signals, or words, which they use to express the meanings. I assume, for simplicity, that the agents can both express and understand these words without difficulty, that is, that the signals can be transmitted without error. The agents also have a dynamic lexicon of associations between words and meanings, which they use both to decide which signals to send, and to decide on an interpretation for the signals they receive. Each entry in the lexicon contains a signal s, a meaning m, a count u of how many times the pair has been used, and a confidence probability p representing the agent's confidence in the association between the signal and meaning, or the proportion of times in which s has been used that it has been associated with m. More formally, $p(s, m)$ can be expressed as

$$p(s, m) = \frac{u(s, m)}{\sum_{i=1}^{l} u(s, i)} \tag{3}$$

where l is the number of entries in the lexicon.[1]

Having successfully undertaken a discrimination game and found a discriminatory meaning, one agent (the *speaker*) utters a signal which represents this meaning. A second agent (the *hearer*) receives the signal together with the original context of objects used by the speaker. The hearer does not know which object was the speaker's target object, but tries despite this to infer the intended meaning solely from the context and from its own previous experiential history, stored in its lexicon as described above. Having inferred a meaning, the hearer then deduces the object to which it thinks the speaker was referring; successful communication occurs when the speaker's original target object is the same object as that which is identified by the hearer's meaning. It is not necessary that the agents use the same agent-internal meaning, only that both agents *refer* to the same object, or pick out the same object in the world. Importantly, neither speaker nor hearer is given *any* feedback on whether the meaning was successfully interpreted.

1 Further details of this communication model and of the structure of the agents' lexicons can be found in [23].

This kind of communicative model, therefore, relies neither on the explicit transfer of meaning nor on feedback to guide the learning. The algorithms for deciding which signal to choose to express a meaning, and for deciding which meaning to interpret a signal as, are therefore crucial to the success of the model. Oliphant and Batali [20] have demonstrated an ideal strategy for achieving an accurate communication system between two agents under these circumstances, which they dub *obverter*. Essentially, this strategy boils down to the speaker choosing signals which it knows the hearer will understand correctly. Unfortunately, true obverter learning assumes that the speaker has access to the lexicons of the other members of the population, so that it can choose the optimal signal for each meaning. Such mind-reading is of course unrealistic, and more damagingly returns us to a telepathic world in which communication using signals is not actually necessary. In order to avoid this, we modify the obverter strategy, by allowing the agent to read only *its own* mind, and using this as a basis for decision making; the speaker therefore chooses the signal that *it itself* would be most likely to understand if it heard the signal in this context.

The hearer, on the other hand, on hearing a signal, has only one source of information apart from the signal itself: the context in which the word was heard. It knows neither the target object to which the speaker is referring, nor the meaning which the speaker has in mind for the signal. The hearer creates a list of *possible meanings*, namely every meaning in its conceptual structure which identifies *any one* of the objects in the context and distinguishes it from all the other objects in the context. The hearer has no reason to prefer any one of these possible meanings over another yet, so each of them is paired with the signal and *lexicalized*, that is, its usage and confidence probabilities in the lexicon are updated. Once all the possible meanings have been lexicalized, the hearer searches through the list of possible meanings, and chooses the one in which it has the highest confidence. If the agent has equally high confidence in more than one meaning, then it chooses one of those meanings at random. The object which this meaning identifies is then compared with the original target object of the speaker's discrimination game, to determine the success of the communicative episode. Neither agent receives any information, however, about the success or failure of the episode.

3.3 Meaning Structure and Communication

Before investigating the interactions between meaning creation and communication, we need to verify that the modified obverter strategy can deliver successful communication without explicit meaning transfer. In order to do this, we therefore temporarily dispense with the meaning creation algorithms, and instead predefine the agents' conceptual systems. Figure 3 shows the communicative success rates for two agents whose meanings have a similarity measure of 80% (left) ($\sigma = 0.8$), and for two agents with identical, synchronized meanings (right) ($\sigma = 1$). The communicative success rate is the proportion of communicative episodes in which the target object described by the speaker is identified by the hearer.

We can immediately see on the right of Figure 3 that when $\sigma = 1$, the communicative success rate rises rapidly from zero, stabilizing as it approaches 1. In principle, the success rate will reach 1, but this is not guaranteed in a particular population over a finite time scale. On the left of Figure 3, we see that when $\sigma = 0.8$, the communicative success rate again rises rapidly in the initial period, and then stabilizes around the level of σ. Given an infinite time scale, we can expect the communicative success rate to equal the agent meaning similarity, and even over a finite time scale it forms a good approximation.

Figure 3 shows very clearly the strong link between the level of meaning similarity and the rate of successful communication. As we have eliminated both explicit meaning

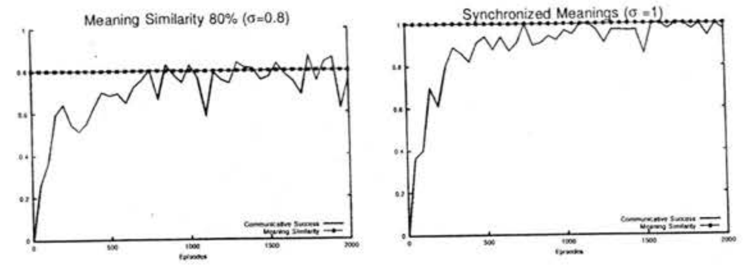


Figure 3. Levels of meaning similarity and communicative success.

transfer and also feedback from the agents to guide their interlocutors to the "correct" answer, unlike models such as those described by Steels and Kaplan [26], we force the agents to infer the meanings of words from the set of possible meanings in each context. It is clear that it is impossible for an agent to attach a word to a meaning which does not exist in its conceptual structure, and so we find inevitably that only those words which refer to shared concepts are successfully used in communication. I have also shown previously [23] how words referring to unshared meanings inevitably suffer semantic drift over time, such that they come to refer to more general meanings which are shared by the agents.

Agents, therefore, can learn communication systems without the explicit transfer of meanings, without knowledge of the topic of conversation, and without feedback about the success of the conversation guiding to the correct meaning. Successful communication arises by the context-driven disambiguation of signals, as long as agents can infer meaning from their experiences in the world. The level of communicative success is very strongly dependent on the level of meaning similarity shared by speaker and hearer.

4 The Standard (or Unbiased) Model

We have seen the importance of synchronized conceptual structure for the development of successful communication without explicit meaning transfer, but how likely is it that synchronization will occur? In this section I investigate the levels of meaning similarity, and by implication communicative success, achieved in a standard, unbiased model. This will also provide a baseline with which to compare the effects of adding cognitive and communicative biases to the agents, as well as external environmental factors such as the structure of the world and the experiences of the agents. The standard model is built on a world with two agents and twenty randomly generated objects. Each object is described in terms of ten features, and each agent has ten corresponding sensory channels on which it can build discrimination trees. The agents play a fixed number of discrimination games, with each agent having an equal probability of being chosen to play the discrimination game. There are five objects in the context, including the target object, unless otherwise stated.

If the size of the context increases, each discrimination game becomes a closer approximation to picking out one individual object from the complete set of objects in the world. An undesirable consequence of this is that the meanings created also identify particular objects in the world. In real human languages, however, words (except possibly some names) do not identify *individuals*, but rather *kinds* [8]. Experiments

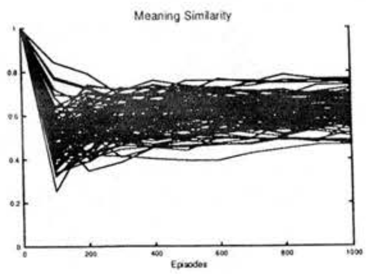


Figure 4. Agent meaning similarity (σ) rates in the standard world. 100 runs overlaid, with each run represented by one line on the graph. The mean ($\bar{\sigma}$) at 1000 episodes is 0.62 (0.61–0.64), with a coefficient of variation of 0.10.

have shown that a level around five provides a suitable balance between developing meanings which identify individuals (with large contexts) and providing the agents with too much information (with small contexts).

Figure 4 shows the level of meaning similarity between the two agents. We can see that overall there is a moderate amount of variation, with no runs producing very high or very low levels of meaning similarity. Meaning similarity is always artificially high at the beginning of each run, because both agents have sensory channels without any tree growth, and therefore identical conceptual structure. As the agents fail in the discrimination tasks, and create new meanings which are not necessarily the same as each other's, overall levels of meaning similarity fall. They then stabilize when the agents have created sufficient conceptual structure to succeed in the discrimination tasks, and there is no further need for much meaning creation. To measure the relative variation we see in Figure 4, I have taken a cutoff point of 1000 episodes, and calculated the average (mean) agent meaning similarity $\bar{\sigma}$ and the coefficient of variation (CoV), which is the standard deviation expressed as a percentage of the mean.[2] I express $\bar{\sigma}$ together with a 95% confidence interval, recognizing that the particular 100 runs of the simulation we have carried out only represent a sample drawn from an infinite set of runs. In the standard model, therefore, we expect to get meaning similarity rates of about 62%, which is not high enough to produce a very successful communication system under normal circumstances. In the following sections, I investigate how variations on this standard model will affect the levels of meaning similarity which the agents achieve.

5 Cognitive Biases and Tree Growth Strategies

In order to explain the apparent paradox of child language acquisition, researchers have regularly appealed to several particular cognitive biases, including the *object bias* [16], which states that a child will assume that an unfamiliar word names a whole object, rather than a particular property of it, and the *shape bias* [14], which states that a child is more likely to assume that an unfamiliar word refers to the shape of an object rather than to other properties such as its color or taste. In our model, the channels are intrinsically meaningless, so we cannot speak in terms of particular properties, but we can investigate how more abstract biases affect the construction of conceptual categories.

2 The standard deviation is scaled relative to the mean so that we can more accurately compare results from distributions with different means.

When a discrimination game fails, the agent chooses a channel on which a node will be refined. This is done on the basis of the channel's bias b_{a_n}, where a identifies the agent and n the number of that agent's sensory channel. The bias is specified when the agent is "born," and does not change during the simulation; it is equivalent to the probability of channel n being chosen for refinement. In the standard model, each channel bias is the same (i.e., there is a uniform bias distribution), and so the agent essentially chooses a channel at random each time, but the channel biases can of course be defined according to particular probability distributions. We will now look at *random* biases, where the bias for each channel is chosen randomly at the start of the simulation; and *proportional* biases, which are defined according to a fixed probability distribution. With proportional bias allocation, the bias on each channel represents a fixed proportion p of the remaining bias, taking into account biases which have already been allocated, as follows:

$$\begin{aligned} &\text{if} \quad n = 0, \qquad b_{a_n} = p \\ &\text{if} \quad n > 0, \qquad b_{a_n} = p\left(1 - \sum_{i=a_0}^{a_{n-1}} b_i\right) \end{aligned} \tag{4}$$

Because the biases represent probabilities, they are always scaled after allocation so that the sum of biases for each agent equals 1. For instance, if p were 0.5, and the agent had five channels, then the biases would be allocated as in Table 2. We can also see that the allocation of biases by proportions is deterministic, so if two agents have the same value of p, then they will have identical cognitive biases. Unless specified otherwise, p is set to 0.5 for all simulations reported here. Under proportional bias allocation, channels with lower numbers always have higher biases, but this is purely an artefact of the implementation, and nothing in the results relies on it.

As well as changing the biases, and therefore the likelihood of tree growth occurring on particular channels, we can also define completely different strategies for the channel choice. In addition to the *probabilistic* method, where the agent chooses a channel at random based on the biases described above, we will investigate another strategy, when the agent searches through its channels in order of their biases, until it finds a refinement which would have resulted in successful discrimination *in this particular discrimination game*, had the refinement already taken place. If no channel which meets this criterion is found, then no refinement takes place.

A crucial feature of this strategy, which I call the *intelligent tree growth strategy*, is that a refinement will always make a helpful distinction in at least the particular discrimination game during which it was created, whereas refinements under the probabilistic strategy are not guaranteed to be successful at all.

Table 3 shows the average rate of agent meaning similarity after 1000 episodes, averaged over 100 runs of the simulations as above, with both the tree growth strategies (probabilistic and intelligent) and the channel bias allocations (uniform, random, and proportional) being varied.[3] Counterintuitively, we find that the best results are achieved under the uniform, standard model which we looked at in Figure 4. The same level is achieved if agents have proportionally allocated biases, suggesting that the important factor is that in both these cases the agents' biases are *identical*. When the agents have random biases, on the other hand, then the level of meaning similarity drops to just over 50%. Under the intelligent strategy, it is interesting that the level of

3 The combination of uniform biases and intelligent tree growth strategy is not included, because the intelligent tree growth strategy is based on searching the channels in order of their probabilities; if these are all equal, then there is no obvious way to order them except randomly, which makes the search equivalent to a random, or probabilistic, choice.

Table 2. Allocation of biases under the fixed proportional method, with $p = 0.5$.

Channel n	Bias b_{a_n}	Scaled Bias
0	0.5	0.5161
1	0.25	0.2581
2	0.125	0.129
3	0.0625	0.0645
4	0.03125	0.0323

Table 3. How different tree growth strategies and cognitive biases affect average agent meaning similarity rates.

Strategy	Biases	$\bar{\sigma}$	CoV
Probabilistic	Uniform	0.62	0.10
	Random	0.52	0.18
	Proportional	0.62	0.18
Intelligent	Random	0.39	0.35
	Proportional	0.43	0.30

meaning similarity is even lower, and the variation very high, with some runs producing meaning structures with almost no similarity at all.

So why do agents produce very divergent conceptual structures when they use the intelligent tree growth strategy? The intelligent strategy always focuses refinements on channels which would have succeeded, and, other things being equal, channels which already have high levels of tree growth are more likely to produce a discriminatory meaning than those which have only very general meanings. Therefore, after a few initial refinements have been made, the intelligent strategy tends to focus further refinements on those channels on which trees have already been grown, and so divergence is therefore almost inevitable under this strategy, unless the initial refinements made by the agents happen to be the same.

6 The Principle of Contrast

Biases which may help explain language acquisition are not just proposed in relation to meaning creation, but also to communication; Clark [6], for instance, proposed the *principle of contrast* (PoC), that every difference in a signal corresponds to some difference in meaning, whereas Markman [17] put forward the closely related ***mutual exclusivity assumption*** (MEA), that children assume that objects do not belong to more than one category. For example, Markman and Wachtel [18] describe how experimenters present children with a banana and a whisk, and then ask them to "show me the fendle." The children tend to interpret *fendle* as referring to the whisk, and it is hypothesized that this is because they already know a word for the banana, so they assume that the unfamiliar word must refer to the unfamiliar object. More recently, these suggestions have been complemented by further research showing how language itself appears, to a certain extent, to shape the learner's meaning structure despite innate biases [15].

The crucial idea underlying both the PoC and the MEA, which can be expressed simply as "every difference in a signal corresponds to some difference in meaning" and implies that there are therefore no true synonyms, can be implemented in our model by ensuring that when an unfamiliar signal is encountered, an agent will create a new meaning which corresponds to one of the objects in the context, and assume that the new signal corresponds to this meaning. This means that meaning creation

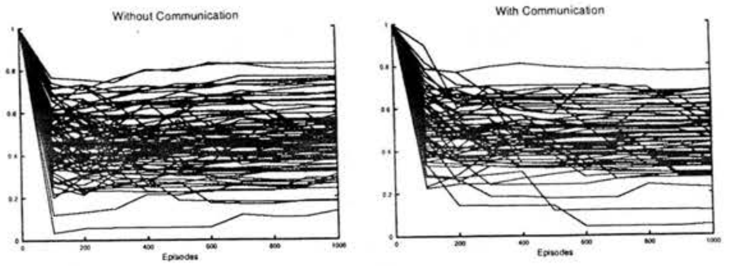


Figure 5. Meaning similarity rates with discrimination-driven meaning creation (left), $\bar{\sigma} = 0.43$ (0.41–0.46), and with the addition of communication-driven meaning creation (right), $\bar{\sigma} = 0.47$ (0.46–0.48).

can therefore now be triggered by two mechanisms in the model: not only failure in the discrimination game, but also failure in the interpretation of an unfamiliar word.

Figure 5 shows how adding meaning creation driven by failure in communication to the model actually has very little effect on the overall level of meaning similarity. We can see that there is a slight increase in $\bar{\sigma}$, but if we use the Kolmogorov-Smirnov (KS) statistic, which expresses how different two distributions are [7], we find that there is no statistical difference between the two sets of results. This would initially appear somewhat surprising, given the frequency with which such heuristics are apparently invoked in the learning of words by children, but in this model it is explained by the fact that the extra information received by the hearer when it receives an unfamiliar word, which it uses to create a new meaning, does not sufficiently help the hearer to build a conceptual structure closer to that of the other agent. Because the words created in the model do not identify individual objects, the occurrence of a new, unfamiliar word is relatively rare. Even when this does occur, the meaning creation process itself is of course unguided, so there is no guarantee that the hearer will build appropriate new conceptual structure, as there is no external pressure to maximize meaning similarity.

7 Environmental Factors

7.1 Experience

This model of empirical meaning creation is based on the agents' building their conceptual structure in response to failures in their interactions with the world, and it would seem reasonable therefore to investigate the importance of the particular situations which they experience. Humans who have similar experiences create distinctions based on those experiences which can be unnoticed or irrelevant to others who have not had them, leading to the creation of particular specialized terminology or jargon to name these distinctions.

In order to investigate how much of the agents' conceptual structure is influenced by the order in which they encounter certain objects and sets of objects, I have implemented simulations in which both agents are given identical discrimination games to perform. Each discrimination game itself still consists of a random target object to be distinguished from a random set of objects, but both agents now undertake the same discrimination game, creating meanings when they fail as in previous experiments. Table 4 shows the levels of meaning similarity achieved when the agents are given identical discrimination games to perform, compared to the results in our reference table (Table 3) when they have different, randomly chosen games. Large values of the

Table 4. How the agents' experiences affect average agent meaning similarity rates.

Strategy	Biases	$\bar{\sigma}$ Diff. exp.	Same exp.
Probabilistic	Uniform	0.62	0.63
	Random	0.52	0.54
	Proportional	0.62	0.64
Intelligent	Random	0.39	0.54*
	Proportional	0.43	1.00**

KS statistic show that the meaning similarity distributions are statistically significantly different; in this article, distributions where $p < 0.05$ are denoted by an asterisk (*), and those where $p < 0.01$ are denoted by a double asterisk (**).

We can clearly see that under the probabilistic strategy, there are no significant differences when the agents have identical experiences, but that in contrast, the intelligent strategy produces significantly increased levels in meaning similarity, under both random and proportional biases. Indeed, if the agents have the same biases and the same experience, we have in effect a deterministic situation, and so it is no surprise that we find complete meaning synchronization ($\sigma = 1$) in this case.

7.2 A Clumpy World

The world in which we live is not uniformly random; indeed, there are many constant properties behind the phenomena we encounter, which can be described in terms of physical and chemical laws. We know, for instance, that unsupported objects will always fall until they reach a lower surface. Scientists can measure the gravitational field which causes this, and we know that its magnitude decreases as the object moves further from the center of the planet; yet in practical terms, the objects in our world do not differ in the gravitational field applying to them. In terms of a space of possible worlds, all the objects in our world are *clumped* together in one section of the space, where the gravitational field is always constant.

Bloom [3] describes how babies use the structure in the world, such as the properties of objects, to make sense of it through categorization and, ultimately, in deciphering the meaning of words. K. Smith [24] has shown how compositional systems are more likely to emerge in generalizing agents when the environment exhibits a high degree of structure. In this model, I investigate how the agents fare in the meaning construction task in a world which is structured or constrained in certain ways, and I explore how the meaning similarity which emerges differs from that in a random world.

In a *clumpy* world, the objects are grouped together in some way and this is implemented in our model by giving each member of a group *identical* feature values for some particular feature (such as the gravitational field applying to them). This means that the objects in a particular group are therefore a priori indistinguishable on this channel, no matter how many times the discrimination tree is refined, and so the objects can only be told apart using meanings created on another sensory channel. In the random world, we could consider each object as a group in itself, with each group containing just one object; in the clumpy world, we choose the number of groups arbitrarily according to the channel and the number of objects in the world. The number of groups on channel c, $g(c)$, is taken as follows:

$$g(c) = \frac{O}{c+1} \tag{5}$$

Table 5. Allocation of groups in a clumpy world.

Channel c	0	1	2	3	4	5	6	7	8	9
Groups $g(c)$	20	10	7	5	4	4	3	3	3	2

Table 6. How the structure of the world affects average agent meaning similarity rates.

		$\bar{\sigma}$	
Strategy	Biases	Random world	Clumpy world
Probabilistic	Uniform	0.62	0.70*
	Random	0.52	0.59*
	Proportional	0.62	0.68*
Intelligent	Random	0.39	0.82**
	Proportional	0.43	0.88**

where O is the number of objects in the world. If there is no exact division, then $g(c)$ is always rounded up to the next whole number.

In a world of 20 objects, therefore, the number of groups on each channel will be as shown in Table 5. We can see that the channels toward the end of the list have few groups, and so are much less likely to be of any use in a discrimination game, though we also note that none is completely useless if all objects fall into one group (this would only happen under this setup if the agents had more sensory channels than there were objects in the world). The groups are arbitrarily biased so that more distinctions can be made on low-numbered sensory channels, just as the proportional allocation of biases was biased toward low-numbered sensory channels. If the structure of the world is biased in a certain direction, it makes sense, if we want to appeal to some selectionist motivation for the existence of the cognitive biases, for the channels to be biased in a similar way.

Table 6 shows that all tree growth strategies produce significantly higher levels of meaning similarity than in simulations under the same conditions in a uniformly random world. The probabilistic strategy produces significantly increased levels of meaning similarity under all conditions where the order of the agents' experiences did not have any significant effect. Under the intelligent strategy, the levels of meaning similarity have more than doubled in comparison with those achieved in the uniformly random world, and the differences are highly statistically significant ($p < 0.01$).

An intelligent meaning creation strategy, therefore, results in poor meaning similarity levels if the agents are in a random world, but it is very good at taking advantage of any structure in the world, and produces very high meaning similarity levels in a clumpy world.

8 Summary

In this article, I have described a model of empirical meaning creation and of the evolution of communication, in which successful communication can emerge without innate meanings and without the explicit transfer of meanings; I have also described the importance of meaning synchronization in the model. Furthermore, I have investigated meaning similarity levels under various conditions, experimenting with various cognitive, communicative, and environmental factors, motivated by research into how children acquire and learn what words mean.

The structure of the world plays a large role in determining which strategy of meaning creation will create a conceptual structure which is most likely to result in successful communication. If the objects in the world are distributed randomly, then the agents can do no better than create meanings based on their innate biases, and reasonably high similarity will occur when the agents happen to have the same biases. If the world is structured, on the other hand, then it is much better for the agents to use an intelligent strategy for meaning creation, which takes account of the structure in the world to a much greater degree.

References

1. Arita, T., & Koyama, Y. (1998). Evolution of linguistic diversity in a simple communication system. *Artificial Life, 4*, 109–124.
2. Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 111–172). Cambridge, UK: Cambridge University Press.
3. Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
4. Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life, 8*, 25–54.
5. Chomsky, N. (1980). *Rules and representations*. London: Basil Blackwell.
6. Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Erlbaum.
7. Conover, W. J. (1999). *Practical nonparametric statistics*, 3rd ed. New York: Wiley.
8. Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.
9. Greenberg, J. H. (1966). *Universals of language*, 2nd ed. Cambridge, MA: MIT Press.
10. Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.
11. Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: The computer simulation of social life* (pp. 157–189). London: UCL Press.
12. Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford, UK: Oxford University Press.
13. Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–203). Cambridge, UK: Cambridge University Press.
14. Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*, 299–321.
15. Levinson, S. C. (2001). Covariation between spatial language and cognition, and its implications for language learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 566–588). Cambridge, UK: Cambridge University Press.
16. Macnamara, J. (1972). The cognitive basis of language learning in infants. *Psychological Review, 79*, 1–13.
17. Markman, E. M. (1989). *Categorization and naming in children: Problems of induction. Learning, development and conceptual change*. Cambridge, MA: MIT Press.
18. Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology, 20*, 121–157.
19. Nehaniv, C. L. (2000). The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In *Proceedings of the AISB Symposium: Starting from Society—The application of social analogies to computational systems, 19–20 April 2000* (pp. 73–84). AISB.

20. Oliphant, M., & Batali, J. (1997). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter, 11.*

21. Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707–784.

22. Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

23. Smith, A. D. M. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen & P. Sosík (Eds.), *Advances in artificial life: Proceedings of the 6th European Conference on Artificial Life,* Lecture Notes in Artificial Intelligence, 2159 (pp. 381–390). Heidelberg: Springer-Verlag.

24. Smith, K. (2002). *Compositionality from culture: The role of environment structure and learning bias* (Technical report). Theoretical and Applied Linguistics, University of Edinburgh.

25. Steels, L. (1996). Perceptually grounded meaning creation. In M. Tokoro (Ed.), *Proceedings of the Second International Conference on MultiAgent Systems, ICMAS-96* (pp. 338–344). Menlo Park, CA: AAAI Press.

26. Steels, L., & Kaplan, F. (2002). Bootstrapping grounded word semantics. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 53–73). Cambridge, UK: Cambridge University Press.

Smith, A. D. M. (2003b). Semantic generalisation and the inference of meaning. in W. Banzhaf, T. Christaller, J. Ziegler, P. Dittrich and J. T. Kim (Eds.), *Advances in Artificial Life: Proceedings of the 7th European Conference on Artificial Life* (pp. 499–506). Heidelberg: Springer-Verlag.

Semantic Generalisation and the Inference of Meaning

Andrew D.M. Smith

Language Evolution and Computation Research Unit,
School of Philosophy, Psychology and Language Sciences, University of Edinburgh
andrew@ling.ed.ac.uk

Abstract. In this paper, a computational model of a successful negotiated communication system is presented, in which language agents develop their own meanings in response to their environment and attempt to infer the meanings of others' utterances. The inherent uncertainty in the process of meaning inference in the system leads to variation in the agents' internal semantic representations, which then itself drives language change in the form of semantic generalisation.

1 Introduction

Modern evolutionary linguistics is primarily occupied with understanding the apparently paradoxical universality and massive diversity found among human languages; much of the recent work within this paradigm uses A-life computational simulation techniques to explore these questions. It is important to note that the evolution of language should not just be considered in a genetic framework, but also in a cultural one; children acquire their language based on the linguistic information produced by those around them. Recently, researchers have shown that the cultural transmission of language itself leads to adaptive pressures which can explain some of the characteristic properties of language [4,7,15]. Cultural explanations of the structure of language mean, importantly, that there may well be less need to fall back on the existence of a somewhat ethereal language acquisition device [11].

Many A-life models of language evolution focus on the emergence of syntactic structure in their agents' language, but fail to explain the origin of the semantics on which their 'emergent' syntax is built. One of the most intriguing universal features of language is the ceaseless and inevitable nature of language change, driven by language variation within speech communities [18]; in this paper I present a model of conceptual development and negotiated communication between agents, in which agents can create meanings individually and still co-ordinate a joint language. I show, moreover, that the inherent uncertainty in the process of meaning inference in such a usage-based model of language [5], leads to variation in the agents' conceptual structures, which itself drives language change in the form of semantic generalisation.

2 Meaning Representation and Creation

A large proportion of recent A-life research into language evolution has been focused on the emergence of syntactic structure, or more specifically compositionality [1,4,7].

W. Banzhaf et al. (Eds.): ECAL 2003, LNAI 2801, pp. 499–506, 2003.

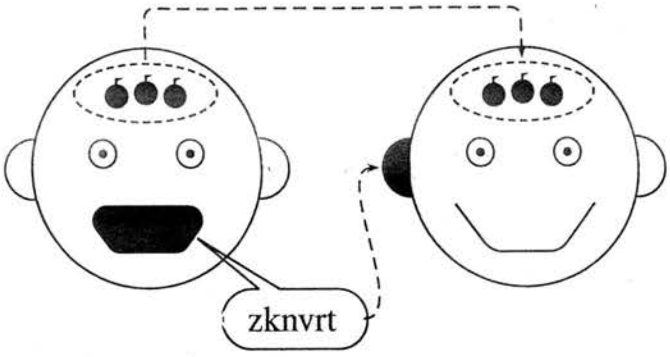


Fig. 1. A communicative episode which consists of the explicit transfer of both a signal 'zknvrt' and a meaning 'three apples' from speaker to hearer.

Although the precise implementations vary to some extent, in all compositionality models structure arises in the signal space as a direct consequence of the agents recognising and coding regularities between parts of signals and parts of meanings. There are a number of problems with models like these, particularly with respect to their semantic representations, which I will discuss here briefly.

Firstly, the agents are always provided with a structured meaning space, which is explicitly linked to an unstructured signal space. As Nehaniv [9] has pointed out, structure develops in the signal space precisely because the agents use signals coupled with already structured meanings; the signals are essentially parasitic on the meanings, and structure only emerges as the agents decipher the pre-defined semantic coding system.

Secondly, no treatment of meaning can avoid addressing the fundamental concepts of both *sense* and *reference* [6]. A basic model of sense relations would include at least, for instance, some notion of antonymy, the relationship between a word and its opposite[1], which is "one of the most important principles governing the structure of languages" [8, p.271], and might be expected also to include other notions such as *hyponymy*, which describes the relationship between a subset and a superset, as for instance between *cat* and *animal*. Reference, on the other hand, is defined in terms of the objects or actions in the external world to which a word refers in a particular context. It is very difficult to relate the 'semantic' structure in such compositionality models to sense relationships in the semantics of real languages in any way, and, even more problematically, there is no reference at all in the meaning systems, either because there is no external world in the experiment at all [1], or because the world is inaccessible to the agents [4,7]. Without reference, and with only a very tenuous link to sense, it is clear that such representations of meaning are actually semantic in name only.

[1] There are many types of opposition in language, including gradable antonyms, such as WET/DRY, expressing meanings on a relative scale; ungradable binary antonyms, such as ALIVE/DEAD, expressing complementary propositions; and converses, such as ABOVE/BELOW, BUT , which refer to the same relationship from opposite points of view.

Thirdly, as I have argued elsewhere [13,14], the lack of reference in the meaning representations of these models leads to an implementation of communication which is seriously flawed, because of its necessary reliance on *explicit meaning transfer* [13]. Figure 1 shows an example of such a communication episode, with the speaker on the left uttering a word *zknvrt*, which is received by the hearer on the right. Simultaneously, however, the meaning THREE APPLES, is transferred directly from speaker to hearer. During communication, the hearer is given explicitly not only the signal, but also both the meaning, and the information that it should make the appropriate association between the particular signal and meaning. Such a model not only sidesteps the very important Quinean [12] problem of how hearers interpret the meanings of unfamiliar words from a set which is in principle infinite, but also actually undermines the need for language at all: if meanings can be transferred accurately by telepathy, then the signals are not being used to convey meaning. There is little motivation, therefore, for the emergence of a system in which the agents spend their time and energy developing a communication system which encodes exactly the same information as another system which they are born with and which already works perfectly.

The semantic model in this paper, on the other hand, tries to avoid these pitfalls; the basic procedure of agent-based grounded meaning creation through discrimination games was originally presented by Steels [16], and has been used by others both in simulated worlds [2,14], and on robots interacting with objects in a real environment [17,19]. An agent is situated in a world made up of abstract objects, with which it interacts by playing discrimination games, which consist of four parts:

Scene-setting: a set of objects, called the context, is chosen from the world and presented to the agent; one of these objects is chosen to be the target of discrimination.
Categorisation: the agent cycles through the objects in the context, returning, for each, an association with one or more of its semantic representations.
discrimination: the agent tries to define a distinctive category for the target object, i.e. a category which both represents the target and does not represent any other object in the context.
Adaptation: the agent modifies its internal conceptual structure in some way.

In the model used in these experiments, the objects in the world have a number of abstract, meaningless features, the values of which are normalised to lie in the range 0.0...1.0, and the agents have a specific sensory channel corresponding to each feature, on which they build a hierarchical semantic representation, or a discrimination tree [16]. Each node on the discrimination tree is a discrete category, corresponding to a particular contiguous segment of the continuous feature value space. Categories are created by splitting the sensitivity range of a node into two discrete, equally sized segments, each of which is therefore sensitive to half the range of the previous node. The trigger for the creation of new meanings is failure in the discrimination game; it is important to note, however, that there is no pre-definition of which categories should be created, nor is there any guarantee that the newly created categories will turn out to be useful in future discrimination games.

Because of this, agents develop different, though equally valid, representations of the same world. I quantify the similarity of two agents' meaning structures by averaging the similarity of the particular discrimination trees built on each of their sensory channels.

If $k(t,u)$ is the number of nodes which trees t and u have in common, and $n(t)$ is the total number of nodes on tree t, then the similarity between any two trees t and u is:

$$\tau(t,u) = \frac{1}{2}\left(\frac{k(t,u)}{n(t)} + \frac{k(t,u)}{n(u)}\right) \tag{1}$$

I then obtain an overall measure of *meaning similarity* σ between two agents, by averaging τ over all their sensory channels. If a_{ij} identifies channel number j on agent i, and each agent has c sensory channels, then the meaning similarity σ between agents a_1 and a_2 is:

$$\sigma(a_1,a_2) = \frac{1}{c}\left(\sum_{j=0}^{c-1} \tau(a_{1j}, a_{2j})\right) \tag{2}$$

If two agents a_1 and a_2 have identical conceptual structures, where $\sigma(a_1,a_2) = 1$, then I refer to their meanings as being *synchronised* [14]. Meanings in this model are expressed using the notation *j-r*, where j identifies the number of the sensory channel, and r is a sequence of 0s and 1s representing the path from the tree's root to the node in question; if the tree's root is on the left, and it branches to the right, then 0 signifies a traversing of the lower branch, and 1 the upper branch.

Importantly, the meanings are grounded in the world, created in response to the environment, and encode both sense and reference relations: the hierarchical nature of the tree means that meanings nearer the root of trees (and therefore with a shorter route r) are more general than those nearer the leaves, and that a node and its daughter nodes can be related by hyponymy; while the meanings also refer to the properties of objects in the world.

3 Communication and the Inference of Meaning

Avoiding the problems with previous models discussed above, the speaker's meaning is *not* transferred to the hearer, in contrast to associative learning models [4,3], nor does the hearer know which object in the context is being referred to, in contrast to Steelsian guessing games [17]; the communication process is made up of three separate sections:

Production: the speaker, having found a distinctive category in a discrimination game, chooses a signal to represent this meaning.

Transfer: the signal is transferred to the hearer.

Interpretation: the hearer interprets the signal from the context in which it is heard.

Each agent maintains a dynamic lexicon of associations between signals and meanings, for use both in production and in interpretation. Each entry in this lexicon contains a signal s, a meaning m, a count u of the pair $< s,m >$'s mutual association, and a representation p of the agent's confidence in the association between the pair $< s,m >$. A signal-meaning pair can be used both by being uttered by the speaker and by being understood by the hearer, so that u is the total number of communicative episodes in which the agent either uttered s to represent m, or interpreted s as representing m. An agent's confidence in a signal-meaning pair is based solely on the relative co-occurrence of signals and meanings, or the proportion of times in which s has been used that it has been associated with m. More formally, $p(s, m)$ can be expressed as:

$$p(s,m) = \frac{u(s,m)}{\sum_{i=1}^{l} u(s,i)}, \tag{3}$$

where l is the number of entries in the lexicon. Communicative success is based on referent identity; speaker and hearer communicate successfully by referring to the same object, but they are not obliged to use the same meaning to do so.

Because meanings are not explicitly transferred between agents, the hearer must infer the meaning of the signal from the context. It is important to note that the hearer does *not* know which object is the target object, and so it must try to come up with descriptions for all the objects in the context. The hearer plays a series of discrimination games and creates a list of possible meanings to consider, each of which describe only one of the objects in the context. The hearer has *no* other information, so all these meanings are equally plausible; it therefore associates each of them with the signal in the lexicon, modifying the values of u and p in the lexical entries accordingly. Having modified its lexicon, it chooses, from the list of possible meanings, the meaning in which it has the highest confidence p. This modification of the lexicon in context is the only way in which the agents learn; in contrast to similar language game models [17], they receive no feedback about the success of either their communication games or their lexicon development. I have shown previously [13] that communication is successful under these circumstances if the speaker chooses a signal that it thinks the hearer will understand. Of course, the speaker does not have access to the hearer's lexicon, as this would defeat the object of ruling out mind-reading, so it bases its decision on what *it itself* would understand, if it heard the signal in this context, without knowing the target object. This technique for choosing signals is a version of the obverter mechanism [10], modified so that the only lexicon an agent can investigate is its own, and which I therefore call ***introspective obverter***.

4 Meaning Similarity and Meaning Variation

The introspective obverter algorithm allows the agents to develop successful communication systems; I have shown elsewhere that to achieve optimal communication, the agents should have synchronised meaning structures [14]. One effect of an optimal communication system, however, is that the agents quickly settle on a common language, which is consistently reinforced through continued use and is completely stable, in contrast to the fluidity of human language.

The development of perfectly synchronised meaning structures is, however, very unlikely, given the inherent randomness in the agents' meaning creation algorithms; so what happens to the agents' language when their meanings are not synchronised and each is trying to communicate their language to the other? I explore this by tabulating detailed extracts from their lexicons through the progress of a simulation. The agents firstly develop most of the meaning structure which enables them to succeed in the discrimination games, and only then do they begin to communicate with each other. Extracts from two sample lexicons after four hundred communicative episodes, showing

Table 1. An extract from two lexicons, showing the high degree of coordination between the signal-meaning pairs.

Agent 1				Agent 2			
Signal	Meaning	Usage	Confidence	Signal	Meaning	Usage	Confidence
bc	3-1	2	0.5	egla	1-11	9	0.45
yq	1-01	8	0.32	bc	3-1	2	0.4
tjop	3-01	7	0.32	tnip	3-10	6	0.35
...	...	...	...	...	...	...	...
egla	1-11	8	0.23	yq	1-01	7	0.25
tnip	3-10	6	0.17	tjop	3-00	6	0.2

Table 2. The meaning exists in both semantic structures, and so context-driven disambiguation leads to an agreement over the meaning of the signal.

Agent 1				Agent 2			
Signal	Meaning	Usage	Confidence	Signal	Meaning	Usage	Confidence
jch	2-000	9	0.69	jch	2-000	9	0.17

the three words in which each agent has the highest confidence, is given in table 1. Although the agents have different levels of confidence in the signal-meaning pairs, they have broadly settled on a common language, with most words referring to the same meaning for both agents.

Later on, the first agent uses a meaning 2-000 in a discrimination game, but has no word for this meaning in its lexicon. These are exactly the circumstances under which we allow lexical innovation to occur, and so it coins a new word jch and utters this to the other agent. The development of the new word jch in both agents' lexicons depends on a number of parameters, including naturally the number of times it is chosen to be uttered, but also crucially whether the meaning to which it is linked is present in both agents' semantic representations, and it is this semantic development of the word jch which I will now discuss.

If the meaning 2-000 does exist in both agents' semantic representations, then we find, after rolling the simulation on a few hundred episodes, that the agents' lexicons contain the extracts shown in table 2. The second agent has associated jch with many different meanings, having heard it in a number of different contexts. The meaning 2-000, however, has occurred in each of these contexts, and so the agent's confidence in this particular signal-meaning pair is higher than in any other; repeated exposure in different contexts has disambiguated the agent's set of possible semantic hypotheses.

If the meaning 2-000 does not exist in both agents' representations, however, then the process of coordination is not so smooth, as we see in table 3. The second agent has again associated jch with a large set of possible meanings, but this time meaning 2-000 cannot be among them, as it is not in this agent's semantic repertoire, and so it is instead most confident in the meaning $4-01$. The agents now each have a different meaning associated with the word, and this has interesting consequences. The second

Table 3. The meaning does not exist in both agents' semantic structure, and no agreement over the meaning of the signal is reached. As a consequence, both agents' confidence in their respective lexical association falls.

Agent 1				Agent 2			
Signal	Meaning	Usage	Confidence	Signal	Meaning	Usage	Confidence
jch	2-000	9	0.47	jch	4-01	6	0.11

agent now uses *jch* to represent meaning 4-01, and utters this in a context in which the first agent's meaning 2-000 is not a possible meaning. This leads the first agent to become less confident in its original association; when agents are using the word for different meanings over time, both agents' confidence in their respective original associations for the word fall. There is now a conflict between the agents' use of the word *jch*, which must in the end be resolved by one agent losing so much confidence in its preferred association that it chooses another meaning for the word.

Because of the hierarchical nature of the meaning creation process, it is more likely that the meanings which the agents have in common, and on which they can agree word associations, are the more general meanings, which are nearer the roots of the discrimination tree. When the conflict over the use of the word *jch* is resolved, therefore, it is likely that the eventual meaning to which it is attached is, other things being equal, a relatively more general meaning than it was originally coined to serve; its use in a variety of contexts by agents who have different semantic structures has led to its meaning becoming more generalised.

The agents' lexicons are of course dynamic, and do not stop developing because the conflict over one word has been temporarily resolved. If the first agent comes across a context in which the original meaning of *jch*, namely 2-000, is needed, it now has no word which it can use and so must coin another one; the generalisation process itself has led to the loss of words from some part of the lexicon, which, if they are needed again, leads inevitably to more innovation. Of course, if meaning 2-000 still does not exist in the other agent's semantic structure, then the same process of conflict over meaning, generalisation and innovation is likely to happen all over again. We find, therefore, that meaning uncertainty, which is inevitable when meanings must be inferred instead of given, leads to the development of a continuous cycle of language innovation and semantic generalisation and extension.

5 Conclusions

I have presented a model in which agents individually create meanings based on their interactions with their external environment, and develop a co-ordinated communication system without either feedback or the explicit transfer of meaning. If the agents have very similar conceptual structures, then they are able to develop optimal communication systems by inferring the meanings of unfamiliar words through their exposure in different contexts. On the other hand, variation in conceptual structure itself creates pressure leading to a cycle of innovation and semantic generalisation. Work is currently under way to explore this trade-off between semantic uniformity which helps communication and semantic variability which drives language change.

References

[1] J. Batali. The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, pages 111–172. Cambridge University Press, Cambridge, 2002.

[2] T. Belpaeme. *Factors influencing the origins of colour categories*. PhD thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel, 2002.

[3] A. Billard and K. Dautenhahn. Experiments in learning by imitation – grounding and use of communication in robotic agents. *Adaptive Behavior*, 7(3/4):415–438, 1999.

[4] H. Brighton. Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54, 2002.

[5] W. Croft. *Explaining Language Change: an evolutionary approach*. Longman Linguistics Library. Pearson, 2000.

[6] G. Frege. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50, 1892.

[7] S. Kirby. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, 5(2):102–110, 2001.

[8] J. Lyons. *Semantics (2 vols.)*. Cambridge University Press, Cambridge, 1977.

[9] C. L. Nehaniv. The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In *Proceedings of the AISB Symposium: Starting from Society — the application of social analogies to computational systems, 19–20 April 2000*, pages 73–84, 2000.

[10] M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, 11(1), 1997.

[11] S. Pinker and P. Bloom. Natural language and natural selection. *Behavioral and Brain Sciences*, 13:707–784, 1990.

[12] W. v. O. Quine. *Word and Object*. MIT Press, 1960.

[13] A. D. M. Smith. Establishing communication systems without explicit meaning transmission. In Jozef Kelemen and Petr Sosík, editors, *Advances in Artificial Life: Proceedings of the 6^{th} European Conference on Artificial Life*, number 2159 in Lecture Notes in Artificial Intelligence, pages 381–390. Springer-Verlag, 2001.

[14] A. D. M. Smith. Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, 9(2):559–574, 2003. in press.

[15] K. Smith. Compositionality from culture: the role of environment structure and learning bias. Paper given at the 4^{th} International Conference on the Evolution of Language, Harvard University, 2002.

[16] L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the international conference on multi-agent systems*. MIT Press, 1996.

[17] L. Steels and F. Kaplan. Bootstrapping grounded word semantics. In E. Briscoe, editor, *Linguistic Evolution through Language Acquisition: formal and computational models*, pages 53–73. Cambridge University Press, Cambridge, 2002.

[18] R. L. Trask. *Historical Linguistics*. Arnold, 1996.

[19] P. Vogt. *Lexicon Grounding on Mobile Robots*. PhD thesis, Vrije Universiteit Brussel, 2000.

Smith, A. D. M. (forthcoming). Mutual exclusivity: Communicative success despite conceptual divergence. in M. Tallerman (Ed.), *Evolutionary prerequisites for language*. Oxford: Oxford University Press.

Mutual Exclusivity: Communicative Success Despite Conceptual Divergence

Andrew D. M. Smith
Language Evolution and Computation Research Unit
Theoretical and Applied Linguistics
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Adam Ferguson Building
40 George Square
Edinburgh
EH8 9LL
United Kingdom
andrew@ling.ed.ac.uk

July 2003

Abstract

In this article, I build on work exploring populations of simulated agents who construct communication systems based on inferring the reference of unfamiliar words from context. In child language acquisition studies, many biases and predispositions have been suggested to explain how the disambiguation of reference appears so unproblematic for children; I explore in particular one of these, the assumption of mutual exclusivity. Previous experiments have shown that, in a model based on meaning inference, communicative success is highly dependent on levels of shared conceptual structure, yet I find that, when the interpretation process is driven by mutual exclusivity, the development of more communicatively relevant conceptual structure is promoted, and communicative success occurs despite conceptual divergence among agents.

Keywords: mutual exclusivity; communicative success; conceptual divergence; meaning inference; meaning creation.

1 Introduction

Traditional explanatory accounts of the evolution of language frequently appeal to a "conventional neo-Darwinian process" (Pinker & Bloom, 1990), assuming that humans have evolved an innate, genetically-encoded language acquisition device, which specifies a formal coding of Universal Grammar (Chomsky, 1965), and which evolved incrementally through a series of steps via natural selection (Jackendoff, 2002). An alternative approach focuses instead on the evolution of linguistic structures themselves, as utterances used and understood by speakers and hearers (Christiansen, 1994; Croft, 2000). Under the latter approach, the continual cycle of expressing and re-interpreting these utterances (Hurford, 2002) drives the cultural evolution of language. Other things being equal, languages which can be readily interpreted and expressed through this cycle are more likely to persist than those which cannot.

An explanation of the evolution of syntactic structure remains the holy grail of evolutionary linguists by researchers in both these traditions, because syntax has been seen as the defining characteristic which separates human language from animal communication systems, and in recent years, computational simulations have been used extensively to shed light on this issue. Kirby (2002), for example, shows that structured signals can develop from unstructured signals through the analysis of signal-meaning pairs and the subsequent generalisation of rules based on the analysis; similar accounts are presented by Batali (2002), whose agents combine and modify phrases based on exemplars of signal-meaning mappings which they receive, and by Brighton (2002), who shows how the poverty of the stimulus is an important factor in the emergence of compositional syntax.

Despite these exciting findings, however, there are some problematic assumptions in models such as these. In particular, the emergence of syntactic structure in the signal space is a direct result of the signals' explicit association with pre-defined meanings (Nehaniv, 2000), and of the explicit transfer of meaning in communication (Smith, 2001). Furthermore, the models often rely on reinforcement learning to guide the learners, although error signals are widely rejected in language acquisition (Bloom, 2000). I have, however, developed a model of meaning creation and communication which addresses these problems and have shown that communication can succeed through the inference of meaning (Smith, 2001, 2003a, 2003b). Crucially, inferential communication allows the development of communication between individuals who do not necessarily share exactly the same internal representations of meaning. This flexibility then opens the possibility of a realistic evolutionary scenario, by allowing both for the necessary variation among individuals, and also for mutations which might enhance the inferential capabilities of one individual, while still allowing them to be communicatively consistent with the rest of the population.

In this paper, I extend my inferential model to explore the usefulness of one of the main psycholinguistic biases proposed to explain how children learn the meaning of words without explicit meaning transfer, Markman (1989)'s *mutual exclusivity assumption*. The remainder of the paper is divided into four parts: In section 2, I describe the signal redundancy paradox which is contained in other models; this pre-determines the outcomes which are achieved and, to a large extent, undermines the strength of their conclusions. In section 3, I focus further on Quine (1960)'s famous problem of the indeterminacy of meaning, and on proposals made by psychologists and psycholinguists to explain how children manage to solve this problem when they acquire language, including, of course, the mutual exclusivity assumption. In section 4, I briefly describe my model of individual, independent meaning creation and negotiated communication which avoids these pitfalls and yet still allows successful communication. I show, crucially, that there is a strong relationship between levels of meaning co-ordination and communicative success. Finally, in section 5, mutual exclusivity is added to the model, and I show that, in contrast to expectations based on my earlier models, this can lead to high levels of communicative success despite agents having divergent conceptual structures.

2 The Signal Redundancy Paradox

Kirby (2002) and Batali (2002), among others, have shown how the simple ability to generalise can result in the emergence of a compositional, 'syntactic' communication system. In their simulations, agents initially create idiosyncratic rules to represent each different meaning they need to express, and each of these rules generates just one signal. Over time, coincidental matches occur between parts of signals and parts of meaning, and the agents create more general rules based on these matches; these rules use symbolic variables and can therefore generate more than one signal. Brighton (2002) shows that if there are pressures on agents which limit their exposure to the language, such as the poverty of the stimulus, then the agents are more likely to encounter general rules than idiosyncratic ones, and so these general rules are preferentially replicated over generations, leading to the eventual evolution of a fully compositional communication system, where the meaning of a signal is made up of a combination of the meanings of its parts and an algorithm for joining these together.

The successful emergence of syntax in such models, however, is completely dependent on the signals being explicitly coupled to meanings which have a pre-defined and complex structure. It is not coincidental that the emergent syntactic structure parallels this semantic structure exactly, as the semantic structure is effectively used as a template against which the syntactic structure is constructed.

2.1 Explicit Meaning Transfer

Figure 1 shows a schematic diagram of the linguistic transfer in such a communicative model, where the utterances are made up of pairs of signals and meanings. We can see that the speaker (on the left of figure 1) utters a signal "*zknvrt*", which is received by the hearer. Simultaneously, the meaning in the speaker's brain, represented in figure 1 by three symbols meant to resemble apples, is transferred directly to the hearer's brain. This explicit linkage of signal and meaning in the communication process means that it is a trivial task for the hearer to learn the association between them.

Models which make this idealisation, therefore, ignore one of the most important and difficult problems facing researchers into the acquisition of language, namely Quine (1960)'s famous problem of the indeterminacy of meaning. Quine presented an imaginary anthropologist, who observes a speaker of an unfamiliar language uttering the word *"gavagai"* while pointing to a rabbit, and then shows that, logically, *"gavagai"* has an infinite number of possible meanings and, moreover, that the collection of further relevant information by the anthropologist will never reduce the number of possible hypotheses which will be consistent with the data; no matter how much evidence is collated, the meaning of *"gavagai"* can never be determined.

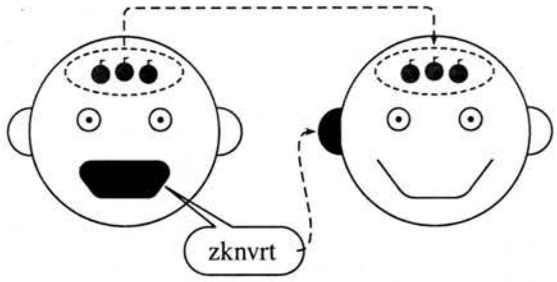


Figure 1: A communicative episode which consists of the explicit transfer of both a signal 'zknvrt' and a meaning 'three apples' from speaker to hearer.

The consequences of the idealisation of the learning process as shown in figure 1 are considerable, not least because if meanings are explicitly and accurately transferable by telepathy, then the signals are not actually being used to convey meaning, and their very role in the model must be called into question; if the agents can transfer meanings between each other, there can be no justification for them to waste time and energy worrying about learning a redundant additional system of signalling. This paradox, which I call the *signal redundancy paradox*, arises whenever meanings are explicitly transferred in communication:

- if the meanings are transferable, then the signals are redundant;
- but if the signals are removed, then to what extent does the model represent communication?

The most obvious way out of this paradox is to conclude that meanings *cannot* be explicitly transferred, but must be inferred from elsewhere.

2.2 Accessibility and Privacy

If there is no explicit meaning transfer, however, how does a hearer know which meaning to associate with a particular signal? The hearer must be able to infer a meaning from somewhere; the most obvious and general source for this is surely the environment in which the agent is placed. This in turn suggests that at least some of the meanings agents talk about are used to *refer* to objects and events which actually happen in the environment. In this way, the agents' meanings are grounded (Harnad, 1990); without the possibility of inferring the signals' reference, real communication cannot emerge. Indeed, the existence of an external world from which meaning can be inferred is crucial to a realistic model of meaning, for without it, any 'meanings' are necessarily abstract and pre-defined. If the meanings do not identify anything in the world, or do

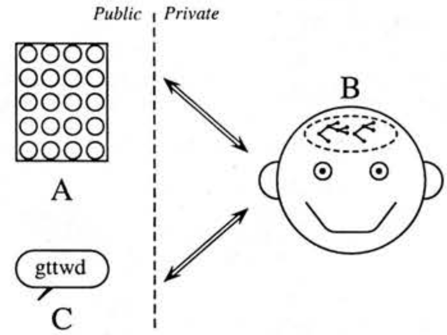


Figure 2: A model of communication which avoids the signal redundancy paradox must have three levels of representation for the agents: an external environment (A); an internal, private semantic representation, represented by the trees in the agent's brain (B); and public signals (C). The mappings between A and B, and between B and C, represented by the arrows, must also be private and inaccessible to other agents.

not have reference, they can only be communicated through explicit transfer, which of course entails the signal redundancy paradox.

In order to avoid the signal redundancy paradox, therefore, there must be at least three levels of representation in the model, as shown in figure 2:

A: an external environment, which is public and accessible to all, which provides the motivation and source for meaning creation;

B: a private, agent-specific internal representation of meaning, which is not perceptible to others;

C: a set of signals, which can be transmitted between agents and is in principle public.

The mere existence of an external world, as for instance in Hutchins and Hazlehurst (1995)'s model of the emergence of a shared vocabulary, is not sufficient to avoid the paradox; if the agents' meanings are publicly accessible, either directly as in Hutchins and Hazlehurst's model where the external scenes *are* the meanings, or indirectly through an accessible mapping between the environment and the meanings, then the signals are again rendered unnecessary. For this reason, note in figure 2 that the mappings between A and B and between B and C fall to the right-hand side of the demarcation line between the public and private domains.

2.3 Inferential Communication

There are at least two possible explanations for how the agents come to have meanings which refer to things in their environment: either the meanings are innate, and have evolved through biological evolution; or they are created by the agents, as a result of their interactions with the environment. Innate meanings are not inherently implausible, but they seem to require either that the number of meanings useful to the agents is small and fixed, or that the world in which the agents exist is very stable and unchanging, so that the evolved meanings which were useful to their ancestors are still of use to the current generation. In practice, then, it is more reasonable to assume that the agents have an innate quality space, as suggested by Quine (1960), within which they create meanings anew in each generation, based on empirical testing of their environment, which allows them to discover which distinctions are communicatively relevant.

Thinking of the communicative function of language as a simple coding system between signals and meanings, however, is problematic not just in terms of the communication model itself, but also in terms of the evolution of such a system. From this perspective, it is important to remember that language is necessarily both reciprocal and cultural. There is no advantage, therefore, in a mutant obtaining a language acquisition device if others do not have one. In addition, however, there is no advantage in *many* mutants having a language acquisition device, if there is no language existing in the community for them to acquire. As Origgi and Sperber (2000) point out, a mutation which allows individuals to infer the meanings of signals can not only provide an explanation for how language got started, through the accidental discovery of what another is referring to, but can also provide a plausible account of the progressive complication of language. For instance, a mutation which promotes the construction of a more complex semantic representation does not, in an inferential model, cause catastrophic effects on communication due to the ensuing mismatch between the speaker's meaning and the hearer's meaning; instead, because communication is based on reference, individuals can have very different internal representations of meanings, and yet still communicate successfully, as I have shown through simulation experiments (Smith, 2003b). Those without the enhanced semantic representation can still communicate with everyone in blissful ignorance, while the mutants might receive an advantage in more accurate or detailed inference of the meaning, and might, in time, develop new symbols to represent the patterns they find in this structure. Indeed, this process of structural development is most obviously attested in historical processes of language change, particularly in the case of grammaticalisation (Hopper & Traugott, 1993), where (more complex) grammatical markers such as case markers and complementisers are created from (less complex) lexical items over generations of inference, a process which has been explicitly described as "context-induced reinterpretation" (Heine & Kuteva, 2002, p.3).

The model I describe, then, departs from previous accounts which assume that language learning is merely equivalent to learning a mapping between signals and predefined meanings. Instead, I argue that it must include at least the construction of empirical meanings, learning which of these meanings are relevant, and learning a mapping between meanings and signals through the inference of meaning in context.

3 Overcoming Indeterminacy

Learning the meanings of words, of course, is utterly unremarkable to children, who effortlessly overcome Quine's problem of indeterminacy; a common suggestion for how this happens is that they are explicitly taught by parents and teachers, by being given feedback on their use of words. Despite the intuitive appeal of this idea, it is actually very rarely observed in practice, and is by no means culturally universal. Lieven (1994), for instance, describes cultures in which parents do not even speak to their children in the initial stages of acquisition, much less provide them with either encouragement or discouragement about their use of words. Bloom (2000), furthermore, describes a study on mute children who clearly could not receive feedback on their own speech, and yet still developed language normally. In view of this, researchers have explored the existence of other constraints within the learners themselves which predispose them to disregard some of the theoretically possible meanings, thus reducing the size of the set of semantic hypotheses, thereby making the set finite and Quine's problem soluble.

Macnamara (1972), for instance, argues that children naturally represent their environment in terms of the objects within it, and that, when learning words, they have a similar *object bias*, under which they automatically assume that a new word refers to a whole object, rather than particular parts or properties thereof. The object bias is indeed a very useful tool in explaining how children might bootstrap language acquisition, but it is not a sufficient explanatory tool for the larger problem, and so many additional biases or restrictions have also been proposed in order to account for more complex facets of word learning. Landau, Smith, and Jones (1988), for instance, discovered experimentally that children are more likely to categorise new objects in terms of their shape, rather than other perceptual features. Markman and Hutchinson (1984) have shown that children categorise objects taxonomically (grouping on the basis of type) rather than thematically (grouping on the basis of relationships between them) when they are learning words, but not otherwise. For instance, when word learning is not involved, a car and a car tyre can be grouped together thematically, but when the car is given a name, and the children asked to find another object which can called by the same name, they are much more likely to find the taxonomically related bicycle.

Interpretation biases, too, have often been proposed; in particular, many of these suggestions, by for instance Barrett (1986), Clark (1987) and Markman (1989), can be summarised by the proposal that "children should construct mutually exclusive extensions of the terms they acquire" (Merriman & Bowman, 1989, p.1). Although there are slight differences between these suggestions in terms of their theoretical and explanatory emphasis, in this paper I will consider them as related versions of an over-arching *mutual exclusivity assumption*. Merriman and Bowman (1989) analyse the implications behind mutual exclusivity, and propose three crucial ways in which the bias could affect the learning of new words; the most important of these, and the only one which does not rely on the explicit naming of objects, is through the *disambiguation of reference*. This phenomenon has been shown experimentally a number of times, particularly by Markman and Wachtel (1988), who describe experiments in which young children were presented with random pairs of objects, one of which is familiar to them, such as a banana or a spoon, and one of which is unfamiliar, such as a lemon wedge presser or a pair of tongs. The children, on being presented with both objects, were asked by the experimenters to "show me the *x*", where *x* was a randomly chosen nonsense syllable. Markman and Wachtel found that the children are much more likely to interpret *x* as referring to the tongs, rather than the banana; they hypothesise that this is because the children already understand a word which means BANANA, and they assume, under the mutual exclusivity bias, that "[w]hen children hear a novel term in the presence of a familiar and unfamiliar object, children are able to use mutual exclusivity to determine the referent of the novel term." (Markman & Wachtel, 1988, p.128). In section 5, I explore how mutual exclusivity can improve the levels of communicative success relative to the shared conceptual structure of agents in my model.

4 Details of the Model

4.1 Independent Meaning Creation

Before investigating the effects of mutual exclusivity, however, it is useful to give a brief description of my basic model of meaning creation and communication, which takes as its starting point the model initially described by Steels (1996). A simple model world is simulated, containing a number of objects, each of which can be described in terms of the values of their observable features. Feature values in the model are real numbers, pseudo-randomly generated in the range [0.0...1.0]; the features themselves, however, are deliberately abstract, with neither specific nor pre-defined meanings, although for ease of understanding, they can of course be considered analogous to features in human language such as 'height', 'smell' or 'colour'. Simulated agents interact with the objects in the world using *sensory channels*; they have the same number of sensory channels as the objects have features, and there is a one-to-one mapping between them. Sensory channels are sensitive to the objects' feature values; specifically, they can detect whether a particular feature value falls between two bounds on a sensory channel. The process of meaning creation takes place through *refinement*, or the splitting of a channel's sensitivity range into two discrete segments of equal size. This results in the formation of two new categories, each sensitive to half the original range. Each category is itself a candidate for further refinement, so producing, over time, a hierarchical, dendritic structure, with the nodes on the tree representing categories, or *meanings* (Steels, 1999). Such structures are shown schematically in the agent's private semantic representation in figure 2.

Interaction with the environment occurs through Steelsian discrimination games, which are made up of the following four constituent parts:

scene-setting: the agent is presented with a specific set of objects, called the *context*, one of which is chosen to be the *target* of discrimination.

categorisation: the agent goes through all the objects in the context, returning for each an association with one or more of its existing semantic representations.

discrimination: the agent tries to find a distinctive category for the target. A category is distinctive if it is a valid representation of the target, and is not a valid representation of any other object in the context.

adaptation: the agent modifies its internal conceptual structure, by refining one of the sensory channels.

Adaptation of an agent's conceptual structure is triggered by failure in a discrimination game. Each agent has a *tree growth strategy* for choosing a channel for refinement, which is based on its cognitive biases and/or the details of the particular discrimination game which failed, as described in Smith (2003b). In a stable world, the agents will eventually always develop a conceptual structure which can succeed in describing every object in the world. Different agents, however, will create different conceptual structures which will each be able to distinguish objects in the world, and so it is useful to be able to measure the level of meaning similarity σ between two agents' conceptual structures (Smith, 2003a).

4.2 Introspective Obverter

Having established that agents can create meanings which are helpful in describing the world around them, I simulate communication without explicit meaning transfer and without feedback by providing the agents with the ability to create simple signals and transmit them without error, and also with a mechanism for associating signals and meanings, an individual dynamic lexicon (Smith, 2003a). In a communication episode, one agent (the speaker) is trying to convey a meaning to another agent (the hearer) by the use of a signal.

Preparatory to communication, a successful discrimination game provides the speaker with a distinctive meaning which has identified the target object from others in the context, and it is this meaning which the speaker tries to convey; it utters a signal to represent the meaning, either taking one from its lexicon, or, if none suitable exists, creating a new one at random. The hearer then tries to infer the meaning of the signal from the context in which it is heard, attempting to deduce which of the objects in the context was identified by the speaker. Successful communication is defined by *referent identity*, which occurs when the object identified by the speaker is the same object as that identified by the hearer. Note that it is not necessary that the agents use the same agent-internal meaning, only that both agents pick out the same object in the world. Importantly, neither speaker nor hearer is given any feedback on whether the communication episode is successful.

This communication model, therefore, relies neither on explicit meaning transfer meaning, nor on feedback guiding learning. The algorithms which determine the agents' behaviour, however, are crucial to its success, and are based on Oliphant and Batali (1997)'s strategy for achieving an accurate communication system in a population of agents, which they dub *obverter*. Essentially, the obverter strategy boils down to the speaker choosing signals which it knows the hearer will understand correctly, rather than choosing signals that it might prefer to say. Unfortunately, true obverter learning in the theoretical situation defined by Oliphant and Batali assumes that the speaker has access to the lexicons of the other members of the population, so that it can choose the optimal signal for each meaning. Such mind-reading is of course unrealistic, and returns us, more damagingly, to a telepathic world and the signal redundancy paradox. In order to maintain the benefits of obverter, whilst also avoiding any reliance on telepathy, I implement a modification to the obverter algorithm, in which I allow the agent to read only its own mind. The agent therefore uses introspection as a basis for decision making, choosing a signal which *it itself* would be most likely to understand if it heard the signal in this context.

Choosing a signal is relatively straightforward, but interpreting that signal is much more difficult; the hearer, to whom this task falls, knows neither the object to which the speaker is referring, nor the meaning which the speaker has in mind for the signal. The hearer creates a list of *possible meanings* or semantic hypotheses, containing every meaning in its conceptual structure which identifies *any one* of the objects in the context and distinguishes it from all the other objects in the context. The hearer has no reason to prefer any one of these possible meanings over another yet, so each of them is paired with the signal in the hearer's lexicon. Having done this for all the possible meanings, the hearer searches through its list of semantic hypotheses, and chooses the meaning m in which it has the highest confidence, which is, as Vogt and Coumans (2003) explain, the highest conditional probability that, given the current signal, the meaning m is expected. The object which the chosen meaning identifies is then compared to the original target object of the speaker's discrimination game to determine the success of the communicative episode.

4.3 Communicative Success

In Smith (2003b), I show that, in such a model, where the agents infer the meanings of words from the contexts in which they hear them, the percentage of successful communicative episodes, or the communicative success rate κ, is highly dependent on the level of conceptual similarity σ between the agents. I experiment with various cognitive biases, environmental factors and meaning creation strategies, to discover the circumstances under which high levels of conceptual similarity are most likely to occur, and show moreover that in a randomly-generated world, the agents cannot improve on creating meanings based on their cognitive biases, using a *probabilistic* tree growth strategy; high levels of conceptual similarity will always arise if the agents share similar values of these biases. In a structured, or clumpy world, on the other hand, then it is much better for the agents to use a more *intelligent*, ecologically rational (Gigerenzer & Todd, 1999) tree growth strategy, which can exploit the information in the environmental structure to a much greater degree.

5 Mutual Exclusivity

Successful communication, therefore, can emerge without the need for innate meanings and without meanings being explicitly transferred between agents, if the agents use introspective obverter to choose signals. On the other hand, communicative success rates are highly correlated with levels of meaning similarity; the exact relationship varies according to the experimental conditions, but it is always a logarithmic curve with communicative success in general slightly higher than meaning similarity. In this section, I implement the mutual exclusivity bias in the model, to see what effects its inclusion has on the development both of coordinated meanings and successful communication.

Two factors, in particular, are crucial in triggering the use of mutual exclusivity, and must be taken into account in developing the model, namely:

signal novelty: the utterance in question is novel, and unfamiliar to the learner;

disambiguation of reference through prior knowledge: the learner reduces the set of meanings under consideration by excluding all objects for which it already understands a word.

Under normal circumstances within my model, the hearer would, on hearing an unfamiliar word in context, build a set of all possible semantic hypotheses and use these to decipher the utterance, as described in section 4. Disambiguating the utterance's reference through prior knowledge, therefore, will allow this set of semantic hypotheses to be reduced; the agent works through the objects in the context, and excludes from consideration all those objects for which it already knows an appropriate word[1]. The agent is then left with a set of unfamiliar objects, and it assumes that the speaker must be referring to one of these objects. The list of semantic hypotheses is therefore based only on these objects, from which the agent interprets the word as before, choosing the meaning in which it has the highest confidence probability.

In addition to this, however, Markman and Wachtel also hypothesise that mutual exclusivity can help the child to develop new meanings, when they cannot interpret an unfamiliar word, because "children would be left with a word for which they have not yet figured out a meaning. This should then motivate children to find a potential meaning for the novel term." (Markman & Wachtel, 1988, p.153). If no interpretation at all is possible, therefore, i.e. there are no appropriate meanings which distinguish any of the unfamiliar objects from all the others in the context, then the agent searches through the unfamiliar objects in turn, trying to create a new, appropriate meaning which will be suitable to describe it in this context. It tests potential refinements on its sensory channels, until it finds a node which, once refined, will distinguish this object from all the other objects in the context, and then creates this new meaning, associating it with the unfamiliar signal.

The hearer's meaning creation process is now very different from the speaker's, both in the mechanism by which it is triggered and in the algorithm through which it is implemented; meaning creation in the hearer now occurs as a result of encountering an unfamiliar word and is a direct attempt to find a relevant interpretation of this word, but in the speaker occurs as a result of failure to discriminate a target object. This implementation of the mutual exclusivity bias differs from my earlier implementation of the principle of contrast (Smith, 2003b); although both sets of simulations use the same framework of meaning creation and communication, in the earlier simulations, the agent did not divide the context into two sets of familiar and unfamiliar words before interpretation, so the list of semantic hypotheses was not reduced, and the meaning creation process was triggered very infrequently.

[1] An appropriate word is defined here as a word which the agent would use, in this context, to describe the object, and which therefore represents a distinctive meaning which would distinguish this object from all the other objects in the current context.

Tree Growth Strategy	Meaning Similarity Mean ($\bar{\sigma}$)	Meaning Similarity CI	Communicative Success Mean ($\bar{\kappa}$)	Communicative Success CI
Probabilistic	0.53	(0.50 – 0.56)	0.70	(0.67 – 0.72)
Intelligent	0.59	(0.56 – 0.63)	0.73	(0.70 – 0.76)

Table 1: Meaning similarity σ and communicative success κ in a randomly-generated world.

5.1 Experimental Results

In the results reported here, the agents in the model each had five sensory channels with cognitive biases distributed uniformly, and the objects in the world were generated randomly. Each simulation consists of 5000 discrimination and communication episodes, and was run 50 separate times, after which the levels of meaning similarity σ and communicative success κ were then averaged, and expressed together with 95% confidence intervals (CI).

Table 1 shows that in this randomly-generated world, for both the probabilistic and intelligent tree growth strategies, the levels of communicative success are indeed slightly higher than those of meaning similarity, as we would expect. On the other hand, we can also see that, in contrast to the experiments in Smith (2003b), there is no significant difference between the tree growth strategies, as their confidence intervals overlap; the large differences I found previously in levels of meaning similarity between these tree growth strategies are almost completely neutralised if the hearer uses mutual exclusivity to guide its interpretation and meaning creation.

Table 2, on the other hand, shows similar experiments in a simulated clumpy world, where the objects are grouped together and given identical values on some features, so that they are a priori indistinguishable on that sensory channel. I showed in a previous article (Smith, 2003b) that the intelligent strategy will produce much higher levels of meaning similarity σ under these circumstances, as it is much more able to exploit the underlying information structure in the environment. We can indeed see in table 2 that meaning similarity is much improved under the intelligent tree growth strategy, as this would predict. Much more interestingly, however, the levels of communicative success in these experiments no longer bear any close relationship with the levels of meaning similarity. We can see that the communicative success levels are very high under both strategies; in particular, even when agents have very dissimilar conceptual structures ($\sigma = 0.35$) under the probabilistic strategy, the use of mutual exclusivity means that the hearer can learn to associate the relevant meanings with the signals and communicate much more successfully than results without mutual exclusivity would suggest.

Agents have different meaning creation processes, which promote very different patterns of conceptual growth. Specifically, the speaker, who creates meaning in response to discrimination game failure, has much more conceptual structure than the hearer, who creates meaning is response to the need to understand unfamiliar words. Moreover, in accordance with Grice (1975)'s conversational principles, the agents use meanings in communication which provide sufficient information to identify the target object, but which are not unnecessarily specific. The meanings which the hearer

Tree Growth Strategy	Meaning Similarity Mean ($\bar{\sigma}$)	CI	Communicative Success Mean ($\bar{\kappa}$)	CI
Probabilistic	0.35	(0.33 – 0.37)	0.81	(0.79 – 0.83)
Intelligent	0.92	(0.88 – 0.95)	0.90	(0.88 – 0.92)

Table 2: Meaning similarity σ and communicative success κ in a clumpy, structured world.

creates under these circumstances are therefore necessarily communicatively relevant, because they can be used to discriminate at least one unfamiliar object from a group of others and therefore describe that object within a communicative episode.

Although the hearer has far fewer meanings, this leads to a situation where those meanings it does have are more relevant and useful for communication, and so the level of communicative success is much higher than might be expected.

6 Summary

I have presented in this paper a model of independent meaning creation and communication, which avoids the signal redundancy paradox and can yet produce successful communication through the inference of meaning from context. The inference of meaning is a crucial factor in the evolution of language, because it can explain both the genesis and the incremental development of negotiated communication systems. Individuals with non-identical semantic representations are able to communicate successfully, while *variation*, necessary to drive language change, and *flexibility*, necessary to allow mutation in semantic representation without catastrophic communication breakdown, both occur naturally as by-products of the meaning inference process itself.

The level of meaning similarity between agents has previously been shown to be very important in predicting the likely level of communicative success in previous simulations. In these experiments, however, the introduction of an assumption of mutual exclusivity into the hearer's interpretation process and the creation of meaning in order to disambiguate the reference of unfamiliar words, leads to the development of fewer, but more relevant meanings in the hearer's conceptual structure, and therefore to relatively high levels of communicative success despite conceptual divergence.

References

Barrett, M. D. (1986). Early semantic representations and early word usage. In S. A. Kuczaj & M. D. Barrett (Eds.), *The acquisition of word meaning* (pp. 39–67). New York: Springer.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 111–172). Cambridge: Cambridge University Press.

Bloom, P. (2000). *How children learn the meanings of words.* MIT Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life, 8*(1), 25–54.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Christiansen, M. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure.* PhD thesis, University of Edinburgh.

Clark, E. V. (1987). The principle of contrast: a constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Erlbaum.

Croft, W. (2000). *Explaining language change: an evolutionary approach.* Pearson.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: the adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 3–34). Oxford: OUP.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York: Academic Press.

Harnad, S. (1990). The symbol grounding problem. *Physica, D 42*, 335–346.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization.* Cambridge University Press.

Hurford, J. R. (2002). Expression/induction models of language evolution: dimensions and issues. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 301–344). Cambridge University Press.

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: the development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: the computer simulation of social life.* London: UCL Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford University Press.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–203). Cambridge: Cambridge University Press.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*, 299–321.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). Cambridge University Press.

Macnamara, J. (1972). The cognitive basis of language learning in infants. *Psychological Review, 79*, 1–13.

Markman, E. M. (1989). *Categorization and naming in children: problems of induction.* Cambridge. MA: MIT Press.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology, 16*, 1–27.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology, 20*, 121–157.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*(3–4).

Nehaniv, C. L. (2000). The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In *Proceedings of the AISB symposium: Starting from society — the application of social analogies to computational systems, 19–20 April 2000* (pp. 73–84).

Oliphant, M., & Batali, J. (1997). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter, 11*(1).

Origgi, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: modularity, language and meta-cognition* (pp. 140–169). Cambridge University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*, 707–784.

Quine, W. v. O. (1960). *Word and object.* MIT Press.

Smith, A. D. M. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen & P. Sosík (Eds.), *Advances in Artificial Life: Proceedings of the* 6^{th} *European Conference on Artificial Life* (pp. 381–390). Springer-Verlag.

Smith, A. D. M. (2003a). *Evolving communication through the inference of meaning.* forthcoming PhD thesis, Philosophy, Psychology and Language Sciences, University of Edinburgh.

Smith, A. D. M. (2003b). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life, 9*(2), 175–190.

Steels, L. (1996). Perceptually grounded meaning creation. In M. Tokoro (Ed.), *Proceedings of the International Conference on Multi-agent Systems.* MIT Press.

Steels, L. (1999). *The Talking Heads experiment* (Vol. I. Words and Meanings). Antwerpen: Laboratorium. (Special pre-edition)

Vogt, P., & Coumans, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation*, *6*(1). (http://jasss.soc.surrey.ac.uk/6/1/4.html)

References

Akhtar, N., & Tomasello, M. (1996). Twenty-four-month-old children learn words for absent objects and actions. *British Journal of Developmental Psychology*, *14*, 79–93.

Allen, K. (2001). *Natural language semantics.* Oxford: Blackwell.

Aristotle. (1933). *The metaphysics* (H. Tredennick, Trans.). London: Heinemann. (Original work published 350 B.C.; Bilingual edition in two volumes)

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*, 163–308.

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, *21*, 547–609.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 875–890.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*, 395–418.

Barrett, M. D. (1986). Early semantic representations and early word usage. In S. A. Kuczaj & M. D. Barrett (Eds.), *The acquisition of word meaning* (pp. 39–67). New York: Springer.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: social and cognitive bases* (pp. 405–426). Cambridge: Cambridge University Press.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 111–172). Cambridge: Cambridge University Press.

Beaken, M. (1996). *The making of language.* Edinburgh: Edinburgh University Press.

Belpaeme, T. (2002). *Factors influencing the origins of colour categories.* PhD thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel.

Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution.* University of California Press.

Bickerton, D. (1990). *Language and species.* Chicago: University of Chicago Press.

Bickerton, D. (1998). Catastrophic evolution: the case for a single step from protolanguage to full human language. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 341–358). Cambridge: Cambridge University Press.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Bloom, P. (2001). Roots of word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 159–181). Cambridge: Cambridge University Press.

Bloom, P. (2002). Mindreading, communication and the learning of names for things. *Mind and Language, 17*, 37–54.

Bowerman, M. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). New York: Basil Blackwell.

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge: Cambridge University Press.

Bowerman, M., & Pederson, E. (1992, December). *Cross-linguistic perspectives on topological spatial relationships.* (Paper presented at the annual meeting of the American Anthropological Association, San Francisco)

Boyer, P. (2000). Evolution of the modern mind and the origins of culture: religious concepts as a limiting case. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language and meta-cognition* (pp. 93–112). Cambridge: Cambridge University Press.

Bresnan, J. (2001). *Lexical-functional syntax.* Oxford: Blackwell.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1), 25–54.

Brighton, H. (2003). *Simplicity as a driving force in linguistic evolution.* PhD thesis, University of Edinburgh.

Brighton, H., Kirby, S., & Smith, K. (2003). Situated cognition and the role of multi-agent models in explaining language structure. In D. Kudenko, E. Alonso, & D. Kazakov (Eds.), *Adaptive agents and multi-agent learning* (pp. 88–109). Berlin: Springer-Verlag.

Briscoe, E. (2002). Grammatical acquisition and linguistic selection. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 255–300). Cambridge: Cambridge University Press.

Brown, P. (1998). Conversational structure and language acquisition: the role of repetition in Tzeltal adult and child speech. *Journal of Linguistic Anthropology*, *8*, 1–25.

Brown, P. (2001). Learning to talk about motion up and down in Tzeltal: is there a language-specific bias for verb learning? In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 512–543). Cambridge: Cambridge University Press.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley.

Budiansky, S. (1998). *If a lion could talk: how animals think.* London: Phœnix.

Burling, R. (1993). Primate calls, human language and non-verbal communication. *Current Anthropology*, *34*, 25–53.

Burling, R. (2000). Comprehension, production and conventionalisation in the origins of language. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language* (pp. 27–39). Cambridge: Cambridge University Press.

Cangelosi, A., & Harnad, S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, *4*(1), 119–144.

Cann, R. (1993). *Formal semantics: an introduction.* Cambridge: Cambridge University Press.

Carey, S. (1978). The child as word-learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality.* Cambridge, MA: MIT Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Carroll, L. (1998). *Through the Looking Glass and What Alice Found There.* Oxford: Oxford University Press. (Original work published 1872; 1998 edition published together with *Alice's Adventures in Wonderland*, introduction and textual notes)

Carstairs-McCarthy, A. (1999). *The origins of complex language.* Oxford: Oxford University Press.

Catford, J. C. (1977). Mountain of tongues: the languages of the Caucasus. *Annual Review of Anthropology*, *6*, 283–314.

Cheney, D., & Seyfarth, R. (1990). *How monkeys see the world: Inside the mind of another species.* Chicago, IL: University of Chicago Press.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: the influence of language-specific lexicalization patterns. *Cognition*, *41*, 83–121.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1988). *Language and problems of knowledge.* Cambridge, MA: MIT Press.

Christiansen, M. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure.* PhD thesis, University of Edinburgh.

Clark, E. V. (1987). The principle of contrast: a constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* London: Erlbaum.

Clark, E. V. (1993). *The lexicon in acquisition.* Cambridge: Cambridge University Press.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Conway, J. H., & Guy, R. K. (1996). *The book of numbers.* Copernicus/Springer.

Corbett, G. G. (1991). *Gender.* Cambridge: Cambridge University Press.

Cowan, J. W. (1997). *The complete lojban language.* The Logical Language Group.

Cowie, F. (1999). *What's within?: nativism reconsidered.* Oxford: Oxford University Press.

Croft, W. (2000). *Explaining language change: an evolutionary approach.* Harlow: Pearson.

Croft, W. (2001). *Radical construction grammar: syntactic theory in typological perspective.* Oxford: Oxford University Press.

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language* (Vol. Volume 2. Phonology, pp. 93–152). Stanford: Stanford University Press.

Cruse, D. A. (1986). *Lexical semantics.* Cambridge: Cambridge University Press.

Curtis, R., & Elton, B. (1987). Ink and incapability. In *Black Adder the Third.* BBC Television [television programme].

de Boer, B. (2001). *The origins of vowel systems.* Oxford: Oxford University Press.

de Boer, B. (2002). Evolving sound systems. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 79–97). Berlin: Springer.

de Jong, E. D. (2000). *Autonomous formation of concepts and communication.* PhD thesis, Vrije Universiteit Brussel.

de León, L. (2001). Finding the richest path: language and cognition in the acquisition of verticality in Tzotzil (Mayan). In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 544–565). Cambridge: Cambridge University Press.

Deacon, T. (1997). *The symbolic species.* London: Penguin.

Dedrick, D. (1998). *Naming the rainbow: colour language, colour science and culture.* Kluwer.

Deuchar, M., & Quay, S. (2000). *Bilingual acquisition: theoretical implications of a case study.* Oxford: Oxford University Press.

Devitt, M. (1981). *Designation.* New York: Columbia University Press.

Dunbar, R. (1996). *Grooming, gossip and the evolution of language.* London: Faber and Faber.

Dupré, J. (1983). *The disorder of things: metaphsyical foundations of the disunity of science.* Cambridge, MA: Harvard University Press.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71–99.

Fasold, R. (1984). *The sociolinguistics of society.* Oxford: Blackwell.

Firchow, I. B., & Firchow, J. (1969). An abbreviated phoneme inventory. *Anthropological Linguistics, 11*(9), 271–276.

Fodor, J. A. (1975). *The language of thought.* Sussex: Harvester Press.

Fodor, J. A. (1998a). *Concepts: where cognitive science went wrong.* Oxford: Clarendon Press.

Fodor, J. A. (1998b). The trouble with psychological Darwinism. *London Review of Books, 20*(2).

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik, 100,* 25–50.

Friedl, J. E. F. (2002). *Mastering regular expressions* (2nd ed.). Cambridge: O'Reilly.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: the adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 3–34). Oxford: Oxford University Press.

Goddard, C. (1998). *Semantic analysis: a practical introduction.* Oxford: Oxford University Press.

Goddard, C., & Wierzbicka, A. (Eds.). (1994). *Semantic and lexical universals: theory and empirical findings.* Amsterdam: John Benjamins.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10,* 447–474.

Goodman, N. (1954). *Fact, fiction and forecast.* London: University of London.

Greenberg, J. H. (1966). *Universals of language* (2nd ed.). Cambridge, MA: MIT Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York: Academic Press.

Harnad, S. (1990). The symbol grounding problem. *Physica, D 42*, 335–346.

Harnad, S. (1996). The origin of words: a psychophysical hypothesis. In B. Velichkovksy & D. Rumbaugh (Eds.), *Communicating meaning: evolution and development of language* (pp. 27–44). London: Erlbaum.

Hashimoto, T. (1997). Usage-based structuralization of relationships between words. In P. Husbands & I. Harvey (Eds.), *Proceedings of the Fourth European Conference on Artificial Life* (pp. 483–492). Cambridge, MA: MIT Press.

Hashimoto, T. (2001). The constructive approach to the dynamical view of language. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 307-324). London: Springer Verlag.

Hauser, M. D. (1996). *The evolution of communication.* Cambridge, MA: MIT Press.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization.* Cambridge: Cambridge University Press.

Hudson, R. A. (1984). *Word grammar.* Oxford: Blackwell.

Hudson, R. A. (1995). *Word meaning.* London: Routledge.

Hurford, J. R. (1974). Exclusive or inclusive disjunction. *Foundations of Language, 11*, 409–411.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua, 77*, 187–222.

Hurford, J. R. (1991). The evolution of critical period for language acquisition. *Cognition, 40*, 159–201.

Hurford, J. R. (1999). Individuals are abstractions. *Behavioral and Brain Sciences, 22,4*, 620–621.

Hurford, J. R. (2000). Social transmission favours linguistic generalization. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 324–352). Cambridge: Cambridge University Press.

Hurford, J. R. (2001). Protothought had no logical names. In J. Trabant & S. Ward (Eds.), *New essays on the origin of language* (pp. 117–130). Berlin: de Gruyter.

Hurford, J. R. (2002). Expression/induction models of language evolution: dimensions and issues. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 301–344). Cambridge: Cambridge University Press.

Hurford, J. R. (2003). The neural basis of predicate argument structure. *Behavioral and Brain Sciences*, *26*(3).

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: the development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: the computer simulation of social life.* London: UCL Press.

Jackendoff, R. (1990). *Semantic structures.* Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Katamba, F. (1989). *An introduction to phonology.* London: Longman.

Keil, F. C. (1992). *Concepts, kinds and cognitive development.* Cambridge, MA: MIT Press.

Kirby, S. (1999). *Function, selection and innateness: the emergence of language universals.* Oxford: Oxford University Press.

Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 303–323). Cambridge: Cambridge University Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, *5(2)*, 102–110.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–203). Cambridge: Cambridge University Press.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language.* London: Springer Verlag.

Kvasnička, V., & Pospíchal, J. (1999). An emergence of coordinated communication in populations of agents. *Artificial Life*, *5*(4), 319–342.

Labov, W. (1973). The boundaries of words and their meanings. In C.-J. N. Bailey & R. W. Shuy (Eds.), *New ways of analysing variation in English* (pp. 340–373). Georgetown University Press.

Lakoff, R. T. (1976). *Language and women's place.* New York: Harper and Row.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Langacker, R. W. (1987). *Foundations of cognitive grammar.* Stanford, CA: Stanford University Press.

Levinson, S. C. (2001). Covariation between spatial language and cognition, and its implications for language learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 566–588). Cambridge: Cambridge University Press.

Liddell, H. G., & Scott, R. (1980). *Greek-English Lexicon (Abridged).* Oxford: Clarendon Press.

Lieberman, P. (1984). *The biology and evolution of language.* Cambridge, MA: Harvard University Press.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). Cambridge: Cambridge University Press.

Lightfoot, D. (1999). *The development of language: acquisition, change, and evolution.* Oxford: Blackwell.

Livingstone, D., & Fyfe, C. (1999). Modelling the evolution of linguistic diversity. In D. Floreano, J. Nicoud, & F. Mondada (Eds.), *Advances in Artificial Life: Proceedings on the Fifth European Conference on Artificial Life* (pp. 704–708). Berlin: Springer.

Locke, J. L. (1993). *The child's path to spoken language.* Cambridge, MA: Harvard University Press.

Lorenz, K. (1966). *Evolution and modification of behaviour.* London: Methuen.

Lyons, J. (1977). *Semantics (2 vols.).* Cambridge: Cambridge University Press.

Macnamara, J. (1972). The cognitive basis of language learning in infants. *Psychological Review*, *79*, 1–13.

Macnamara, J. (1982). *Names for things: a study of human learning.* Cambridge, MA: MIT Press.

MacWhinney, B. (1985). Hungarian language acquisition as an exemplification of a general model of grammatical development. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 1069–1155). New Jersey: Erlbaum.

Maggio, R. (1989). *The non-sexist word finder: a dictionary of gender-free usage.* Oryx.

Markman, E. M. (1989). *Categorization and naming in children: problems of induction.* Cambridge. MA: MIT Press.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology*, *16*, 1–27.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, *20*, 121–157.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*, 813–815.

Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution.* Oxford: Oxford University Press.

Merriman, W. E. (1986). Some reasons for the occurrence and eventual correction of children's naming errors. *Child Development*, *57*, 942–952.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3–4).

Mitchell, M. (1996). *An introduction to genetic algorithms.* Cambridge, MA: MIT Press.

Morgan, J. L., & Travis, L. L. (1989). Limits on negative information in language input. *Journal of Child Language*, *16*, 531–552.

Morton, E. S., & Page, J. (1992). *Animal talk.* New York: Random House.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. R. Shanks (Eds.), *Knowledge, concepts and categories.* Cambridge, MA: MIT Press.

Nehaniv, C. L. (2000). The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In *Proceedings of the AISB symposium: Starting from society — the application of social analogies to computational systems, 19–20 April 2000* (pp. 73–84). AISB.

Nettle, D. (1999). *Linguistic diversity.* Oxford: Oxford University Press.

Newport, E. H., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition.* Cambridge University Press.

Ninov, V., Gregorich, K. E., Loveland, W., Ghiorso, A., Hoffman, D. C., Lee, D. M., Nitsche, H., Swiatecki, W. J., Kirbach, U. W., Lane, C. A., Adams, J. L., Patin, J. B., Shaughnessy, D. A., Strellis, D. A., & Wilk, P. A. (2002). Editorial note: Observation of superheavy nuclei produced in the reaction of ^{86}Kr with ^{208}Pb. *Physical Review Letters*, *89*(3).

Nowak, M. A., Plotkin, J. B., & Jansen, V. A. A. (2000). The evolution of syntactic communication. *Nature*, *404*, 495–498.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, *7*(3/4), 371–384.

Oliphant, M., & Batali, J. (1997). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, *11*(1).

Origgi, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: modularity, language and meta-cognition* (pp. 140–169). Cambridge: Cambridge University Press.

Orr, M. J. L. (1996). *Introduction to radial basis function networks* (Tech. Rep.). Centre for Cognitive Science, University of Edinburgh.

Pauwels, A. (1999). Feminist language planning: has it been worthwhile? *Linguistik Online*, *2*.

Pawley, A. (2001). Some problems of describing linguistic and ecological knowledge. In L. Maffi (Ed.), *On biocultural diversity: linking language, knowledge and the environment* (pp. 228–247). Washington: Smithsonian Institute Press.

Peirce, C. S. (1897/1955). Logic as semiotic: The theory of signs. In J. Buchler (Ed.), *The philosophical writings of peirce* (pp. 98–119). New York: Dover Books.

Pinker, S. (1994). *The language instinct.* London: Penguin.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707–784.

Polani, D., Uthmann, T., & Dautenhahn, K. (2000). Evolution of sensors in nature, hardware and simulation. *Künstliche Intelligenz, 1,* 33–35.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review, 19*(1-2).

Putnam, H. (1975). The meaning of "Meaning". In K. Gunderson (Ed.), *Language, mind and knowledge* (pp. 131–193). Minneapolis: University of Minnesota Press.

Quine, W. v. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Quine, W. v. O. (1969). Natural kinds. In *Ontological relativity and other essays* (Vol. I, pp. 114–138). New York: Columbia University.

Regier, T. (1995). A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics, 6*(1), 63–88.

Robinson, A. (2002). *Lost languages: the enigma of the world's undeciphered scripts.* McGraw-Hill Professional.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (p. 111-144). New York: Academic Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Sag, I. A., & Wasow, T. (1999). *Syntactic theory: a formal introduction.* Stanford, CA: CSLI Publications.

Sampson, G. (1997). *Educating Eve: The 'Language Instinct' debate.* London: Cassell.

Saussure, F. d. (1916). *Cours de linguistique générale.* Paris: Payot.

Schlesinger, I. M. (1979). Cognitive and linguistic structures: the case of the instrumental. *Journal of Linguistics, 15,* 307–324.

Schmaus, A. (1961). *Lehrbuch der Serbokroatischen Sprache.* München: Max Hüber Verlag.

Schoenemann, P. T. (1999). Syntax as an emergent characteristic of the evolution of semantic complexity. *Mind and Machines*, *9*, 309–346.

Schwartz, S. P. (1980). Natural kinds and nominal kinds. *Mind*, *89*, 182–195.

Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). *Oxford English Dictionary* (2nd ed.). Oxford: Clarendon Press. (http://oed.com)

Slobin, D. I. (1973). Cognitive prerequisites for the acquisition of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 173–208). New York: Holt, Rinehart and Winston.

Smith, A. D. M. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen & P. Sosík (Eds.), *Advances in Artificial Life: Proceedings of the* 6^{th} *European Conference on Artificial Life* (pp. 381–390). Heidelberg: Springer-Verlag.

Smith, A. D. M. (2003a). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, *9*(2), 175–190.

Smith, A. D. M. (2003b). Semantic generalisation and the inference of meaning. In W. Banzhaf, T. Christaller, J. Ziegler, P. Dittrich, & J. T. Kim (Eds.), *Advances in Artificial Life: Proceedings of the* 7^{th} *European Conference on Artificial Life* (pp. 499–506). Heidelberg: Springer-Verlag.

Smith, A. D. M. (forthcoming). Mutual exclusivity: Communicative success despite conceptual divergence. In M. Tallerman (Ed.), *Evolutionary prerequisites for language.* Oxford: Oxford University Press.

Smith, K. (2001). The importance of rapid cultural convergence in the evolution of learned symbolic communication. In J. Kelemen & P. Sosik (Eds.), *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life* (pp. 637–640). Heidelberg: Springer-Verlag.

Smith, K. (2002a). *Compositionality from culture: the role of environment structure and learning bias* (Tech. Rep.). Language Evolution and Computation Research Unit.

Smith, K. (2002b). The cultural evolution of communication in a population of neural networks. *Connection Science*, *14*(1), 65–84.

Smith, K. (2003). *The Transmission of Language: models of biological and cultural evolution.* PhD thesis, University of Edinburgh.

Smith, L. B. (1999). Children's noun learning: how general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language* (pp. 277–303). New York: Erlbaum.

Smith, L. B. (2001). How domain-general processes may create domain-specific biases. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 101–131). Cambridge: Cambridge University Press.

Snyman, J. W. (1970). *An introduction to the !Xũ (!Kung) language.* Cape Town: Balkema.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meanings: object terms and substance terms. *Cognition, 38*, 179–211.

Song, J. J. (2001). *Linguistic typology: Morphology and syntax.* Harlow: Longman.

Spelke, E. S. (1994). Initial knowledge: six suggestions. *Cognition, 50*, 443–447.

Steedman, M. (2000). *The syntactic process.* Cambridge, MA: MIT Press.

Steels, L. (1996a). Emergent adaptive lexicons. In P. Maes, M. Mataric, J.-A. Meyer, J. Pollock, & S. Wilson (Eds.), *From animals to animats 4: proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour.* Cambridge, MA: MIT Press.

Steels, L. (1996b). Perceptually grounded meaning creation. In M. Tokoro (Ed.), *Proceedings of the International Conference on Multi-agent Systems.* Cambridge, MA: MIT Press.

Steels, L. (1996c). The spontaneous self-organization of an adaptive language. In S. Muggleton (Ed.), *Machine intelligence 15.* Oxford: Oxford University Press.

Steels, L. (1997). Constructing and sharing perceptual distinctions. In M. van Someren & G. Widmer (Eds.), *Proceedings of the European Conference on Machine Learning.* Springer-Verlag.

Steels, L. (1999). *The Talking Heads experiment* (Vol. I. Words and Meanings). Antwerpen: Laboratorium. (Special pre-edition)

Steels, L., & Kaplan, F. (2002). Bootstrapping grounded word semantics. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: formal and computational models* (pp. 53–73). Cambridge: Cambridge University Press.

Talmy, L. (2000). *Towards a cognitive semantics (2 vols.).* Cambridge, MA: MIT Press.

Tomasello, M. (1999). *The cultural origins of human cognition.* Harvard: Harvard University Press.

Tomasello, M. (2001a). The item-based nature of children's early syntactic development. In M. Tomasello & E. Bates (Eds.), *Language development: the essential readings* (pp. 169–186). Oxford: Blackwell.

Tomasello, M. (2001b). Perceiving intentions and learning words. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 132–158). Cambridge: Cambridge University Press.

Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology, 30*, 639–650.

Tomasello, M., & Farrar, J. (1986). Joint attention and early language. *Child Development, 57*, 1454–1463.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4*, 197–212.

Traill, A. (1985). *Phonetic and phonological studies of !XÓÕ Bushman.* Hamburg: Helmut Buske Verlag.

Trask, R. L. (1993). *A dictionary of grammatical terms in linguistics.* London: Routledge.

Trask, R. L. (1996). *Historical linguistics.* London: Arnold.

Turkel, W. J. (2002). The learning guided evolution of natural language. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 235–254). Cambridge: Cambridge University Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.

van Turennout, M., Ellmore, T., & Martin, A. (2000). Long-lasting cortical plasticity in the object naming system. *Nature, 3*(12), 1329–1334.

Van Valin, R., & LaPolla, R. (1997). *Syntax: Structure, meaning and function.* Cambridge: Cambridge University Press.

Vaux, B., & Pəsiypa, Z. (1997). The C°əžə dialect of Abkhaz. *Harvard Working Papers in Linguistics, 6.*

Vihman, M. M. (1996). *Phonological development: the origins of language in the child.* Oxford: Blackwell.

Vogt, P. (2000). *Lexicon grounding on mobile robots.* PhD thesis, Vrije Universiteit Brussel.

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research Journal, 3*(3), 429–457.

Vogt, P., & Coumans, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation, 6*(1). (http://jasss.soc.surrey.ac.uk/6/1/4.html)

Vygotsky, L. S. (1986). *Thought and language.* Cambridge, MA: MIT Press. (Original work published 1934)

Ward, C. R., Gobet, F., & Kendall, G. (2001). Evolving collective behavior in an artificial ecology. *Artificial Life, 7*(2), 191–209.

Wexler, K. (1991). On the argument from the poverty of the stimulus. In A. Kasher (Ed.), *The Chomskyan turn.* Oxford: Blackwell.

Whorf, B. L. (1956). *Language. thought and reality.* Cambridge, MA: MIT Press.

Wierzbicka, A. (1996). *Semantics, primes and universals.* Oxford: Oxford University Press.

Wierzbicka, A. (1997). *Understanding cultures through their key words: English, Russian, Polish, German and Japanese.* Oxford: Oxford University Press.

Wittgenstein, L. (1953). *Philosophical investigations.* Oxford: Blackwell.

Wittgenstein, L. (2001). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuiness, Trans.). London: Routledge. (Reprinted from *Annalen der Naturphilosophie*, 1921; original title *Logisch-philosophische Abhandlung*)

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication, 18*, 47–67.

Wray, A. (2000). Holistic utterances in protolanguage. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: social function and the origins of linguistic form* (pp. 285–302). Cambridge: Cambridge University Press.

Yamauchi, H. (2001). The difficulty of the Baldwinian account of linguistic innateness. In J. Kelemen & P. Sosík (Eds.), *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life* (pp. 391–400). Heidelberg: Springer-Verlag.

Ziegler, J., & Banzhaf, W. (2001). Evolving control mechanisms for a robot. *Artificial Life*, *7*(2), 171–190.

Zuidema, W. H. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS '02).* Cambridge, MA: MIT Press.

Index of Authors

Index of Languages

Index of Subjects